Bacterial Diversification Through the Evolution of Cross-Feeding Interactions

Dissertation

zur

Erlangung der Naturwissenschaftlichen Doktorwürde

(Dr. sc. mat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Magdalena San Roman Rincon

aus

Uruguay

Promotionskommission

Prof. Dr. Andreas Wagner (Vorsitz)

Prof. Dr. Jordi Bascompte

Prof. Dr. Hanna Kokko

Prof. Dr. Owen Petchey

Zürich, 2020

Acknowledgments

In 2015, I moved to Switzerland to experience science in a top University, to work with a Professor I admired and excited about the thought of learning German. Five years later, I feel lucky and grateful for the opportunity I had, and proud of all I have accomplished. But none would have been the same if I had not encountered great people on the way. Here, I thank you.

Thank you Andreas for receiving me in your group. I have sincerely enjoyed working by your side. I truly appreciate the scientific freedom you gave me, and how you always made a slot in your busy schedule to deal instantly with my things, such as papers or the thesis. I thank you for your patience, enthusiasm and knowledge. You taught me to express my thoughts more clearly, and everything I know about writing. I hope these five years were just the first years of a long collaboration.

In addition to my advisor, I want to thank the rest of my committee: Jordi Bascompte, Hanna Kokko and Owen Petchey. You motivated me at every meeting with insightful questions and comments. The ecological flavor that this thesis took is partly due to you. Thank you also for your time and kind words.

I would like to express my most sincere gratitude to Matias Arim and Sara Mitri for reviewing my thesis. I truly appreciate the time you invested in me.

Thank you Luis Acerenza for everything you taught me in my initial steps in science, and for encouraging me to take this path.

I also want to thank Annette and Francesco for all these years of secretarial and IT help.

Thank you Michael for translating the summary of this thesis into German. Thank you Josh for correcting the English of the introduction of this thesis.

I would also like to thank all the students with whom I worked in the courses BIO134, BIO334, BIO351 all these years. I thank Cheng, Jeruscha, and especially James. Without knowing it, these students gave me energy and motivation.

What I value most from my time in the Wagner lab is what I have learned with the Wagnerians in coffee breaks, at lunch, out for beers, in the ASVZ or at hiking trips. Thank you Maca, Mariana, Manuel, Charles, Ilaria, Jordi, Michael, Tess, Felix, Alex, Jose, Heidee, Josh, Debbie, Eugenio, Carla, Pouria, Jia, Ali, Kassia, Rzgar, Pierre, Fahad, Gabriel, Camille, Yolanda, Kathleen, Athena, Miriam, Shraddha, José, Sinisa, Ljiljiana, Bharat, Amir, Caua, Andrei and Diego. Thanks to you, I felt excited

about going to the lab every morning.

I want to thank a few Wagnerians more extensively. Maca, you are the most thoughtful and considerate friend. You remember every birthday and small event in my life, and inform me of every conference or seminar I should attend. Tomati, Francesc, you, and every family member to come, represent my family in Zürich. Thank you Josh for all those days when we shared the most profound conversations about life, and for all those other days when we shared the loudest laughter. Thank you Ila for your joy and friendship. You stayed in the group only briefly, but it was enough to infect us with your great personality. Jordi, funny and smart. You have been a role model to me. Miriam and Shraddha. So many coffees, and walks to the post office or material center. So many chats. Three different backgrounds and cultures, and so much we share and agree about how to live our lives. Whenever I take some time to reflect about my friendship with any of you, I am surprised to realize we have only shared a few years together. I feel I have known you all my life.

Also, thank you Lu for your support. Though you were distant, I have always felt you close. Thank you Zermat. With you I got the perfect balance of Latin and German culture.

Most importantly, I want to thank my family. Both the one in Uruguay and the one I made here in Europe. Thank you to my grandparents, siblings and Sofi. You gave me happiness with every call, video and message. Specially, I want to thank my parents. You supported my decision to move far away because you understood it was a good experience for me, though it was difficult for you. Also, thank you both for teaching me, from a very young age, to persevere and work hard, but to always put family, friends and health first.

Last, I want to thank Ugaitz. By your side everything is enjoyable and easy. I want to thank you for moving to Zurich. You left your home, family and friends to join my Swiss adventure. I want to thank you especially for making these last months of long working hours painless.

Thank you.

Vielen Dank.

Eskerrik asko.

Gracias.

Summary

The interplay between evolutionary and ecological processes results in the extraordinary diversity of life we observe on Earth. Understanding these processes is one of the fundamental aims of biology. Species interactions are one of the multiple factors thought to influence the evolution and persistence of organisms. An example of the importance of species interactions comes from experimental work with the bacterium *Escherichia coli*, in which an isogenic population grown in a homogeneous glucose medium evolved into two strains whose coexistence is maintained through cross-feeding interactions: one strain consumes glucose from the medium and excretes a by-product metabolite on which the other strain thrives.

In this thesis, I investigate the role of cross-feeding on the evolution and persistence of microbial communities. I do so theoretically by modeling species with metabolic networks. Specifically, I use genome-scale metabolic networks containing all the metabolic reactions known to take place in a given organism; random viable metabolic networks which are networks similar in complexity to genome-scale networks, but that have an otherwise random complement of reactions; and methods developed for their analysis like Flux Balance Analysis. Metabolic modeling is convenient for my purpose since it allows me to infer, rather than assume, interactions between organisms such as cross-feeding.

In chapter 2, I perform an exhaustive search of the cross-feeding potential of different organisms. I begin by exploring the cross-feeding interactions that could arise in populations of *E. coli* grown in glucose medium. To this end, I search for those metabolites that *E. coli* excretes as by-products of growth on glucose and which, in turn, permit growth of those individuals who consume these by-products as sole sources of carbon. Then I change the carbon source of the growth medium from glucose to each one among 180 metabolites, and search for potential cross-feeding interactions in each medium. I find that 58 cross-feeding interactions could arise when *E. coli* grows in glucose medium, and this number rises to almost 10000 when considering all growth media analyzed. I also explore the cross-feeding potential in metabolic systems other than *E. coli*. Specifically, I analyze *Bacillus subtilis*, *Saccharomyces cerevisiae*, and 500 random viable metabolic networks. Similar results emerge from the analysis of these organisms and from random viable networks. Together, these results suggest that bacterial diversity can increase extensively as cross-feeding interactions evolve between individuals in a population.

Whereas experiments demonstrated cross-feeding interactions only for two metabolites - acetate and glycerol - in *E. coli* my computational analysis predicts that 58 metabolites can be cross-fed in the same conditions. In chapter 3, I investigate the reason for this difference. I hypothesize that greater metabolic changes (and more mutations) might be required for the evolution of those computationally predicted cross-feeding interactions that were not experimentally observed. To explore this possibility, I develop a method to find the minimal metabolic rearrangement required to evolve from an ancestral-like to a cross-feeding strain. I find that the metabolic changes required for the evolution of acetate and glycerol cross-feeding are no less complex than those needed for the evolution of eighteen other cross-feeding interactions, and conclude that many other cross-feeding interactions may await discovery.

In species-rich communities, few ecological niches may be free for new species to colonize, which may constrain community biodiversity. At the same time, however, in species-rich communities species-species interactions such as cross-feeding might be frequent, which may promote biodiversity. In chapter 4, I explore how niche availability and niche construction through cross-feeding may affect community biodiversity. To this end, I simulate the assembly of thousands of communities in a chemostat-like environment where species compete for available resources and create new niches as they excrete by-products of metabolism. I observe that because of cross-feeding, communities can accommodate multiple species in environments that initially contained only a single resource. Competition further fosters diversity by reducing the breadth of a species' realized ecological niche. However, once communities consist of dozens of species, no new niches are created, existing niches become filled, these niches cannot be partitioned further, and species diversity reaches a maximum.

Altogether, the results of this thesis reinforce the importance of species-species interactions in the evolution and persistence of communities, and suggest that considerable microbial biodiversity may have evolved and persist due to cross-feeding interactions.

Zusammenfassung

Das Zusammenspiel von evolutionären und ökologischen Prozessen führt zu der außergewöhnlichen Biodiversität des Lebens, die wir auf der Erde beobachten. Das Verständnis dieser Prozesse ist eines der grundlegenden Ziele der Biologie. Interspezifische Wechselwirkungen sind einer der zahlreichen Faktoren, von denen angenommen wird, dass sie die Evolution und das Überleben der Organismen beeinflussen. Ein Beispiel für die Relevanz von interspezifischen Wechselwirkungen stammt aus einem Experiment mit dem Bakterium *Escherichia coli*, in dem sich aus einer anfänglich genetisch identischen Population, die in einem homogenen glukosehaltigem Nährboden wuchs, sich zwei Stämmen entwickelten, deren Koexistenz durch Kreuzfütterung gesichert wurde: Ein Stamm nahm Glukose direkt aus dem Nährboden auf und schied einen Abfallstoff aus, der dem anderen Stamm als Nahrung diente.

In dieser theoretischen Studie untersuche ich die Rolle der Kreuzfütterung für die Evolution und Erhaltung von Gemeinschaften von Mikroorganismen. Ich verwende zur Simulation von unterschiedlichen Arten sogenannte metabolische Netzwerke. Genau genommen verwende ich 'genome-scale' metabolische Netzwerke, die alle von einem bestimmten Organismus bekannten Reaktionen enthalten; nach dem Zufallsprinzip generierte metabolische Netzwerke, die ähnliche Komplexität wie 'genome-scale' Netzwerke aufweisen, aber ansonsten zufällige Reaktionen beinhalten; und Methoden, die für ihre Analyse entwickelt wurden, wie zum Beispiel 'Flux Balance Analysis'. Die metabolische Simulation ist für meine Zielsetzung praktisch, da ich damit direkt nach Kreuzfütterung zwischen Arten suchen kann, anstatt sie a priori anzunehmen.

In Kapitel 2 führe ich eine umfassende Suche nach dem Kreuzfütterungspotential von verschiedenen Organismen durch. Ich beginne mit der Erforschung der Kreuzfütterung, die in Populationen von *E. coli* entstehen könnte, welche auf einem Glukosemedium wachsen. Zu diesem Zweck suche ich nach Metaboliten, welche *E. coli* als Nebenprodukte des Wachstums auf Glukose ausscheidet und die daraufhin von anderen Organismen aufgenommen werden und deren Wachstum möglich machen. Auch untersuche ich neben Glukose noch 180 andere Kohlenstoffquellen auf ihr Kreuzfütterungspotenzial. Ich fand dabei heraus, dass auf Glukose 58 Wechselwirkungen möglich sind, und fand insgesamt fast 10000 mögliche Wechselwirkungen. Außerdem suche ich auch nach dem Kreuzfütterungspotential in anderen metabolischen Netzwerken als dem von *E. coli* (ich analysiere *Bacillus subtilis*, *Saccharomyces cerevisiae* und 500 zufällig generierte Netzwerke). Dabei ergeben sich ähnliche Resultate. Insgesamt

weisen diese Befunde darauf hin, dass Kreuzfütterung ein hohes Potenzial für die Diversifizierung von Bakterien darstellt.

Während Experimente nur für zwei Metaboliten - Azetat und Glycerin - Kreuzfütterung aufzeigten, sagt meine Computeranalyse vorher, dass unter denselben Bedingungen 58 Metaboliten ausgetauscht werden können. In Kapitel 3 untersuche ich den Grund für diesen Unterschied. Ich gehe dabei der Hypothese nach, dass für von mir vorhergesagten Kreuzfütterungen, die nicht experimentell nachgewiesen sind, eine drastischere Veränderung in einem metabolischen Netzwerk notwendig ist (also mehr Mutationen). Um diese Möglichkeit zu untersuchen, entwickle ich eine Methode, um die minimalen metabolischen Veränderungen zu finden, die erforderlich sind um von einem ursprünglich nicht kreuzfütternden Stamm einen kreuzfütternden Stamm zu evolvieren. Ich finde, dass die metabolischen Veränderungen, die für die Entwicklung der Azetat- und Glycerin-Kreuzfütterung erforderlich sind, nicht weniger komplex sind als diejenigen, die für die Evolution von achtzehn anderen Kreuzfütterungen erforderlich sind, und schließe daraus, dass viele andere Kreuzfütterungswechselwirkungen möglicherweise erst entdeckt werden müssen.

In artenreichen Gemeinschaften sind möglicherweise nur wenige ökologische Nischen frei, in denen sich neue Arten ansiedeln können, was zur Einschränkung der Artenvielfalt einer Gemeinschaft führen könnte. Gleichzeitig aber könnte in artenreichen Gemeinschaften Kreuzfütterung häufig sein, was die Biodiversität fördern kann. In Kapitel 4 untersuche ich, wie die Verfügbarkeit von Nischen und der Aufbau von Nischen durch Kreuzfütterung die Biodiversität einer Gemeinschaft beeinflussen können. Zu diesem Zweck simuliere ich den Aufbau von tausenden von Gemeinschaften in einer chemostatähnlichen Umgebung, in der Arten um verfügbare Ressourcen konkurrieren und Nischen schaffen, wenn sie Nebenprodukte des Stoffwechsels ausscheiden. Ich beobachte, dass als Folge der Kreuzfütterung sich diverse Gemeinschaften in einer Umgeben formen können, die ursprünglich nur eine Kohlenstoffquelle enthielt. Des weiteren wird die Biodiversität außerdem noch vom Wettkampf um einzelne Kohlenstoffquellen gefördert, da Wettkampf die Breite einer realisierten ökologischen Nische verringert. Sobald jedoch Gemeinschaften aus Dutzenden von Arten bestehen, werden keine neuen Nischen mehr geschaffen, bestehende Nischen füllen sich, diese Nischen können nicht weiter aufgeteilt werden und die Biodiversität erreicht ein Maximum.

Insgesamt bestätigen meine Resultate die Bedeutung von interspezifischen Wechselwirkungen für die Evolution und das Überleben von Gemeinschaften. Sie legen nahe, dass Kreuzfütterung eine wichtige Rolle im Ursprung und Erhaltung der mikrobiellen Biodiversität spielt.

List of Publications

This thesis is based on the following papers:

a. **M. San Roman** and A. Wagner. (2018) An enormous potential for niche construction through bacterial cross-feeding in a homogeneous environment. *Plos Computational Biology*

b. **M. San Roman** and A. Wagner. (2020) Acetate and glycerol are not uniquely suited for the evolution of cross-feeding in *E. coli*. *Submitted*

c. **M. San Roman** and A. Wagner. (2020) Diversity-creating species interactions can shift ecological diversity limits during community assembly. *To be submitted*

Contents

Chapter 1

Introduction

1.1 Biological diversity

Life's diversity is astonishing. Approximately 8,7 million eukaryotic species (Mora et al., 2011) and a trillion microbial species (Locey and Lennon, 2016) are estimated to inhabit planet earth. Diversity at local scales is no less surprising. More than 300 tree species are found on a single hectare of tropical rainforest (Gentry, 1988), 580 species of benthic invertebrates are found in the deep waters of Lake Baikal in Russia (Sanders, 1968), hundreds of coexisting phytoplankton species are found in marine ecosystems (Cermeño et al., 2014), and between 2000 to 18000 different microbial genomes have been estimated to exist per gram of soil (Daniel, 2005).

In the following subsections I first briefly describe how new diversity is created by different modes of speciation. I also demonstrate the role of niche construction in diversification. Throughout the section I discuss the similarities or differences between populations of microbes and macro-organisms.

1.1.1 Generation of new diversity

Speciation is the evolutionary process by which a population diversifies into two distinct biological species. Before I describe how speciation occurs, I want to note that more than twenty species concepts have been proposed for eukaryotes, based on different properties (De Queiroz, 2007). For instance, the ecological species concept emphasizes the occupation of distinct ecological niches by different species (Andersson, 1990), where a niche is defined as all components of the environment that characterize the requirements of an individual (Begon et al., 2006; Hutchinson, 1957). In contrast, one version of the phylogenetic species concept emphasizes monophyly (Mishler, 1985), in which case a species consists of an ancestor and all of its descendants. Here, I refer to species as in the popular biological species concept of Mayr (Mayr, 1999) which states that species are interbreeding groups separated from other such groups by reproductive barriers (de Queiroz, 2005).

Speciation is commonly divided into four modes based on the geographical separation of individuals

in a population. These modes are allopatric, peripatric, parapatric, and sympatric speciation. Below, I briefly describe the two most contrasting modes, which are allopatric and sympatric speciation.

Most biological diversity probably arose through allopatric speciation (Coyne and Orr, 2004) in which a population gets subdivided by a physical barrier that prevents contact between individuals in the two subpopulations. In this speciation scenario, as mutations arise in each subpopulation, they stay there, because the physical barrier prevents the flow of organisms between subpopulations. With time different mutations accumulate on each side of the barrier. Eventually, the two subpopulations become so different that they can no longer reproduce with one another and a new species is born. A beautiful example of this process involves the snapping shrimp of Central America (Coyne, 2010). The Isthmus of Panama closed gradually some 3.1 million years ago, separating the Caribbean from the Pacific waters. This separation caused multiple speciation events of the snapping shrimp, a crustacean with a distinctive claw that can grow to be half the size of its entire body length and which, when closed, produces a loud snap that gives the snapping shrimp its name. The similarities between snapping shrimp's species on the sides of the Isthmus accompany the closure of the Isthmus, such that more similar species are found in shallower waters which were separated more recently.

In contrast to allopatric speciation, diversity can also arise without any physical barrier through a process called sympatric speciation. This speciation mode can rapidly give rise to different plant species through polyploidy (Wood et al., 2009; Vallejo-Marín et al., 2015), in which case the doubling of chromosome number results in reproductive isolation between progeny and ancestral population. Sympatric speciation is also often driven by ecological factors. For instance, when apples were introduced to North America, individuals of the apple maggot flies changed their plant hosts from hawthorn to apples. Today, the apple maggot fly population feeds on hawthorns or on apples, but not both. This ecological barrier can over time lead to the creation of new fly species, just like a physical barrier would (Feder et al., 1988).

In the context of prokaryotes, controversy is not exclusively centered on the different modes that can create new species but on the concept of species itself (Rosselló-Mora and Amann, 2001). The role that physical barriers play in microbial diversification is not clear and is intensely debated (Whitfield, 2005). Bacterial diversification may be understood in light of ecological speciation, which is the formation of new ecologically distinct populations (i.e., ecotypes) and which can occur either in allopatry or sympatry (Kassen and Rainey, 2004; Van den Bergh et al., 2018). A little over half a century ago, Atwood, Schneider, and Ryan (Atwood et al., 1951) demonstrated that in simple environments composed of single niches (here: resources) the phenomenon of periodic selection is expected, where diversity within a microbial population is purged recurrently by natural selection. In contrast, in complex environments composed of many niches, selection favors the evolution of ecotypes that specialize on each of the available niches (Kassen and Rainey, 2004). Although complex environments have been

regarded as the most important factor for diversification to occur, fluctuating environments have also been shown to facilitate the maintenance or rise of diversity (Cooper and Lenski, 2010; Yi and Dean, 2013; Sandberg et al., 2017).

In simple environments, the growth of one ecotype can create new niches, and thus allow an increase in diversity in much the same way as occurs in complex environments (Rozen and Lenski, 2000a; Rosenzweig et al., 1994). In the subsection that follows I describe the processes by which microbes can transform a simple environment into a complex one and show how this transformation can result in an increase of diversity.

1.1.2 Niche construction and its role in microbial diversification

Organisms may help create niches for other organisms, for example because they serve as hosts to parasites, or because they change the environment in ways that others may profit from. Some organisms modify the environment in a very noticeable way, such as spiders when they create webs, rabbits when they dig burrows, birds when they build nests or beavers when they construct dams. Others do so in more subtle ways, such as plants and earthworms that alter the chemical composition and humidity of soil, or microalgae from Antarctica that absorb light within ice and thus reduce its strength (Jones et al., 1996). Bacteria also change their environment (McNally and Brown, 2015). They consume resources, form biofilms and stromatolites, secrete toxins, metabolites, antibiotics and antibiotic-degrading enzymes.

There are various terms in the literature used to refer to these environmental changes caused by organisms. The differences are subtle. If the environmental change consists of a physical change that provides a niche for another species, then it is referred to as ecosystem engineering (Jones et al., 1996). If the change alters also the species' own niche, then it is called niche construction (Day et al., 2003). And when the modification persists for many generations, the term ecological inheritance is used (Odling-Smee et al., 2003). It has been argued that all these terms are special cases of the extended phenotype concept presented by Dawkins (Dawkins, 2004; Erwin, 2005). In this thesis I use the term niche construction in a broad sense to refer to any environmental change created by an organism, such that this change has direct or indirect consequences on the same organism and on other coexisting organisms.

Cross-feeding

The term cross-feeding was first coined by the British biologist Hermann Reinheimer in 1921 to refer to the exchange of metabolites between organisms belonging to different kingdoms of life. He proposed the term in-feeding for metabolic exchange within a kingdom (D'Souza et al., 2018). The distinction did not catch on and today we use cross-feeding to refer to any exchange of metabolites, regardless of

the kingdom to which producer and receiver of the metabolites belong.

The results of a meta-analysis (D'Souza et al., 2018) based on cross-feeding interactions found in the literature showed that cross-feeding is common between bacteria and between bacteria and members of other kingdoms. There are striking differences in the nature of the metabolites exchanged, depending on the kingdom to which the interacting organisms belong. Bacteria are a great source of nitrogen for protists, fungi and plants, and they receive carbon sources from plants, protists, fungi and other bacteria. Bacteria also exchange multiple amino-acids with animals, and produce vitamins that are taken up by animals or other bacteria.

Cross-feeding interactions play essential roles in the maintenance of communities (Baran et al., 2015; Magnúsdóttir et al., 2017; Goldford et al., 2018; Enke et al., 2019; Blasche et al., 2019). It has also been argued that cross-feeding could contribute to the problem of bacterial unculturability, where bacteria that are observed under the microscope resist cultivation on agar plates (Pande and Kost, 2017). The explanation is that resources lacking in the culture medium are normally supplied by cross-feeding partners.

Cross-feeding can occur with or without physical contact between interacting organisms (D'Souza et al., 2018). Cross-feeding with physical contact can be mediated by vesicles or nanotubes and causes no change in the environment. In contrast, when cross-feeding occurs in a contact-free way, metabolites are first excreted to the environment, thus changing it. Then other organisms may profit by importing these metabolites. This import can, in turn, indirectly affect the producer positively, for example if the released metabolite is toxic for the producer, or negatively, if the recipient benefits extensively from the cross-fed metabolites, thus increasing its population size and competing more strongly with the producer for other resources. Contact-free cross-feeding can be considered a special case of niche construction.

The evolution of cross-feeding interactions facilitates bacterial diversification

The competitive exclusion principle (Hardin, 1960) posits that if two species are in direct competition for a single resource, one of the species eventually goes extinct, and only the species with the larger fitness on the limiting resource survives. In apparent contradiction to this principle, the evolution of stable coexistence of two or more strains has been repeatedly observed in long term evolution experiments performed in homogeneous environments containing glucose as the sole source of carbon. In the following paragraphs I present two independent studies where such coexistence was observed, and explain this apparent contradiction.

The first study was performed on isogenic populations of *E. coli* growing in a glucose-limited chemostat (Helling et al., 1987; Rosenzweig et al., 1994; Treves et al., 1998; Kinnersley et al., 2009, 2014). A chemostat is a culturing vessel experiencing a constant influx of new media and a constant

efflux of culture medium containing microbial cells, unused nutrients, and metabolic excretions. In 11 out of 15 parallel laboratory evolution experiments in such a chemostat, initially isogenic populations of *E. coli* developed genetic and phenotypic polymorphisms (Helling et al., 1987). These polymorphisms were not transient and persisted over hundreds of generations. Further analysis of a polymorphism that evolved in one of these experiments showed that the evolved strains stably coexisted in the chemostat as a result of cross-feeding (Rosenzweig et al., 1994). That is, one strain consumed the primary carbon source glucose and excreted a secondary carbon source (acetate or glycerol), whereas the other strain fed on the secondary carbon source.

These phenotypic differences resulted from a set of mutations in global regulators, as well as in transcription factors and proteins that directly alter glucose transport, central metabolism, fermentative pathways, the TCA cycle and the glyoxylate shunt. Since cross-feeding interactions involving acetate and glycerol evolved, mutations more closely related to the metabolism of these metabolites are of particular interest. An example of such mutation is a cis-regulatory mutation that affects the expression of acetyl CoA synthetase, an enzyme that catalyzes the transformation of acetate to acetyl CoA, which enters the tricarboxylic acid cycle to produce energy. In addition, a structural mutation occurred in the glycerol-3-phosphate repressor, which may result in constitutive expression of glycerol utilization genes (Treves et al., 1998; Kinnersley et al., 2014).

The second study comes from a famous long term evolution experiment led by Richard Lenski which began more than 30 years ago (Pennisi, 2013). In this study, isogenic populations of *E. coli* were grown in glucose limited media, but in batch culture instead of in a chemostat. In contrast to a chemostat, cells in batch culture are exposed to changing conditions as nutrients get depleted and as waste and cells accumulate. In this experiment, isogenic populations evolved into strains that coexisted for at least 10000 generations in nine out of twelve populations (Good et al., 2017). Though it remains unknown what maintains this polymorphism in eight of these populations, the polymorphism of one population (designated Ara-2) has been fully characterized.

In the Ara-2 population, two strains originated after 6500 generations and coexisted for over 12000 generations (Rozen et al., 2005). Their coexistence was also maintained through glucose-acetate cross-feeding (Rozen and Lenski, 2000b; Le Gac et al., 2012; Plucain et al., 2014). Sequencing of clones from each strain sampled at generation 6500 revealed that the population had become hypermutable, with each clone showing on average 199 mutations (Plucain et al., 2014). Transcriptional profiling of clones of both strains sampled at 6500, 17000 and 40000 generations (Le Gac et al., 2012) showed a dramatic change in global expression, with more than 2000 genes changing expression in the evolved relative to the ancestral strain. In addition, the number of genes differentially expressed between the cross-feeding strains kept increasing with time. Many genes more highly expressed in the acetate consumer strain were involved in the Entner-Doudoroff pathway and in the glyoxylate cycle. From

the list of genes more highly expressed in the acetate producer strain, a secondary glucose transporter stands out which might explain the larger glucose consumption observed in this strain.

Taken together, these experiments suggest that *E. coli* can easily diversify in homogeneous environments. It does so by filling niches initially absent from the environment that are created by the organism itself.

1.1.3 Processes involved in the maintenance of diversity

While evolution brings new species into communities, it does not by itself explain how diversity is maintained once established. Ecological factors must be responsible for this maintenance, because a new species will be able to evolve and persist only if it finds its place in a habitat or community.

The number of different species that a community hosts (the community richness) is influenced by the specific order and timing in which species happen to join the community (Chase, 2003; Fukami, 2015), by abiotic factors such as pH, temperature, climate, geography, availability of resources, and environmental heterogeneity, as well as by biotic (species-species) interactions that include competition, predation, mutualism, commensalism and parasitism (Hanson et al., 2012).

Ten years ago, Mark Vellend and Anurag Agrawal proposed a theory of ecological communities (Vellend and Agrawal, 2010) that groups all these low-level processes which influence the presence of an organism into four high-level processes (figure 1.1). Their aim was to facilitate the study and interpretation of the mechanisms of community assembly and biogeographic patterns. In analogy to the four central processes of population genetics theory (selection within species, drift, gene flow, and mutation) the proposed theory is based on four processes - selection among species, drift, dispersal, and diversification. Diversification and dispersal bring organisms into communities, while drift and selection influence the organisms' persistence over time (Vellend and Agrawal, 2010; Vellend, 2016).

Although this theory was not originally conceptualized for the description of microbial communities, the same basic processes guide the assembly of such communities (Hanson et al., 2012; Nemergut et al., 2013). In the next paragraphs I describe these processes in the context of microbes.

Selection, for example, is an important force shaping microbial community assembly (Martiny et al., 2006; Lozupone and Knight, 2007; Chaffron et al., 2010). Though there is extensive evidence for the role of selection via abiotic factors, we know much less about how biotic factors shape microbial communities, since these interactions are much more difficult to study in microbes than in macroorganisms (Nemergut et al., 2013).

Studies on macroorganisms demonstrated that drift is most important when selection is weak, diversity is low and the total number of community members is small (Chase and Myers, 2011). These conditions are also met in some microbial communities (Nemergut et al., 2013) which suggests that drift is an important process influencing the assembly of microbial communities as well.

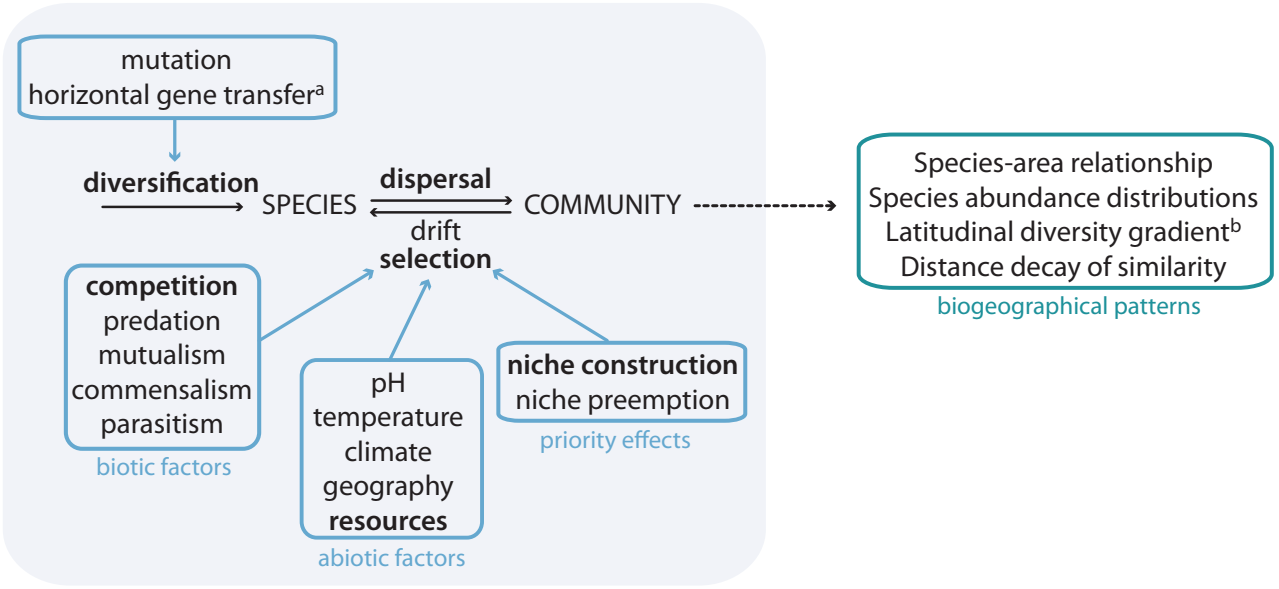


Figure 1.1: Processes involved in the evolution and persistence of diversity. Communities result from four (high-level) processes: selection, drift, dispersal and diversification. In turn, these processes are affected by low-level processes (shown in blue boxes). Biogeographic patterns emerge as a consequence of these processes. The low-level processes and biogeographic patterns shown are only examples and do not pretend to enumerate all possibilities. Superscripts a and b indicate processes or patterns specific to microbes and macroorganisms respectively. The gray box highlights the area of study in this thesis and the processes written in bold are processes I explore specifically.

Last, the genetic diversity of a community can increase as mutations or horizontal gene transfer takes place in or between community members. Another source of community diversity, which is specific to the microbial world, involves dormancy, where a cell can persist for a long time until favorable conditions are met (Wood et al., 2013). A cell might leave the dormant state in a community different from the one it knew when entering dormancy. Additionally, community diversity can increase as organisms disperse and arrive to a new community. Like macroorganisms, microorganisms disperse, but microbial dispersal is typically a passive process. Long-distance transport usually occurs via wind, water or by hitchhiking onto macro-organisms (Fenchel and Finlay, 2004).

The joined actions of selection, drift, dispersal, and diversification gave rise to the biogeographical patterns observed in nature. For instance, in macroorganisms there are latitudinal gradients - diversity increases from the poles to the tropics (Pianka, 1966; Hillebrand, 2004). Various not mutually exclusive hypotheses have been proposed to explain such diversity gradients. For example, the evolutionary speed hypothesis proposes that for a community to have high species richness it must have a long uninterrupted evolutionary history which might be easier to achieve in the tropics where climatic disasters like glaciations are rare. The geographic area hypothesis suggests that the tropics support more species simply because they cover a greater area, which allows for higher population

sizes. Larger population sizes, in turn, reduce the risk of extinction. Competition theory suggests that rich communities, such as those of the tropics, experience greater interspecific competition, which causes a reduction of niche breadth, which allows communities to accommodate more species. All these hypotheses can be reinterpreted in terms of the four high-level processes proposed by Vellend and Agrawal. While the evolutionary speed hypothesis explains the latitudinal gradients in terms of diversification, the geographical area hypothesis and competition theory explain this observation stressing the actions of drift and selection respectively.

Latitudinal gradients seem to be absent in microbes (Fenchel and Finlay, 2004). However, other biogeographical patterns are shared between microbes and macroorganisms. Like communities of macroorganisms, most microbial communities harbor many rare species and just a few common species, a pattern that is known by the name of species abundance distribution (Martiny et al., 2006). Another pattern observed both for microbes and macroorganisms is the distance-decay relationship, in which similarity in species composition between two communities decreases with the geographic distance that separates them (Astorga et al., 2012).

Analysis of massive sequencing data of microbial communities sampled on a wide range of ecological settings has been immensely valuable to understand microbial organization (Chaffron et al., 2010; The Human Microbiome Project Consortium, 2012; Sunagawa et al., 2015; The Earth Microbiome Project Consortium et al., 2017). However, a mechanistic understanding of the process of community assembly is still lacking. Such an understanding would help predict microbial community composition and function, as well as explain the biogeographic patterns observed in nature in a bottom-up manner. This knowledge could also be used to build synthetic communities with potential applications in food and energy production or smart waste treatments (Widder et al., 2016).

1.2 Metabolic modeling

All life is sustained by a complex network of chemical reactions known as metabolism. These reactions comprise enzymatic and transport processes. They allow a cell to get resources and energy from the environment, which is then used for the synthesis of all building blocks required for cell maintenance and growth. Evolution has created organisms with metabolic pathways and regulatory rules encoded in their genomes, which allow them to respond to environmental changes, and survive in different environments.

In this section I review theoretical methods used for the analysis of metabolism. I begin with a brief summary of the history of metabolic modeling and then focus on contemporary methods based on reaction stoichiometry.

1.2.1 History of metabolic modeling and analysis

The study of metabolism originated in the study of its components - metabolic enzymes. Enzymes are catalysts that help convert a chemical reaction's substrates into products by lowering the free energy of activation of the reaction. They can speed up a reaction up to 10 million times (Lehninger et al., 2013), while remaining unchanged by the reaction. They are also highly specific, usually catalyzing the reaction of only one particular substrate or closely related substrates (Keener and Sneyd, 2009).

Although the word enzyme was first coined by Wilhelm Kühne in 1877, it was not until 1897 that Eduard Buchner gave it the meaning we recognize today. Buchner found that sugar was fermented by yeast extracts which contained the necessary enzymes for sugar fermentation, even in the absence of living yeast cells. This discovery yielded him the Nobel Prize in Chemistry in 1907.

In 1913, Leonor Michaelis and Maud Menten proposed the well-known Michaelis-Menten model of enzyme kinetics. The model was originally proposed to explain the enzymatic reaction that catalyzes the hydrolysis of sucrose into glucose and fructose (Johnson and Goody, 2011). Later it was noticed that biochemical reactions involving a single substrate often follow Michaelis–Menten kinetics. The Michaelis-Menten model predicts a hyperbolic relationship between the rate of product formation v and the substrate concentration $[S]$:

$$v = V_{max}\frac{[S]}{K_M + [S]}$$

Here, V_{max} corresponds to the maximum product formation rate and is defined as $V_{max} = k_{cat}[E]_0$ where k_{cat} is the catalytic rate constant and $[E]_0$ the initial (total) enzyme concentration. K_M is known as the Michaelis-Menten constant. It is commonly used to characterize enzymes, and corresponds to the substrate concentration at which product formation is half maximal. It is defined as $K_M = (k_r + k_{cat})/k_f$, where k_f and k_r are, respectively, the forward and reverse rate constants for the formation and breakdown of the substrate-enzyme complex.

However, not all enzymes obey Michaelis-Menten kinetics. For instance, cooperative enzymes and many allosterically regulated enzymes exhibit nonhyperbolic kinetics (Lehninger et al., 2013; Cornish-Bowden, 1979). Cooperativity occurs when an enzyme binds more than one substrate molecule and the binding of one molecule affects the binding of subsequent ones. The reaction rate of cooperative enzymes shows a sigmoidal shape when plotted as a function of substrate concentration (Keener and Sneyd, 2009). Allosteric regulation occurs when molecules (known as regulators) bind to sites on the enzyme that are different from the active site causing an increase or decrease of enzyme activity (Koshland et al., 1966; Changeux, 2011; Keener and Sneyd, 2009). In this way enzymes can be regulated by complicated positive and negative feedbacks which allow precise control of their reaction rates. Various models that predict the kinetic activity of cooperative and allosteric enzymes have been developed (Cornish-Bowden, 1979).

Once models that correctly predicted enzyme activity were available, the study of metabolism

evolved from the study of isolated enzymes to the analysis of metabolic pathways. The nomenclature changed accordingly. For example, the term reaction rate was increasingly replaced by flux rate - the amount of substrate that a metabolic reaction in a pathway converts to product per unit time. And the term flux distribution appeared. It refers to the set of fluxes through every reaction in a pathway.

In 1972, Goldbeter and Lefever (Goldbeter and Lefever, 1972) presented a model to study glycolysis that combined enzyme kinetics and ordinary differential equations. Their simplified model involved the activation of phosphofructokinase (PFK1, an enzyme that catalyzes the third step of glycolysis) by one of its products via an allosteric mechanism. This model could reproduce the oscillatory behavior observed under special conditions in the concentration of metabolites involved in yeast glycolysis (Betz and Chance, 1965; Hess and Boiteux, 1968).

With time, the complexity of the modeled pathways increased. For example, in 1995 a mathematical analysis of the energy and redox metabolism of human erythrocytes was published (Schuster and Holzhütter, 1995), which included the activity of 19 enzymes. A few years later, an analysis which included 30 enzymes was performed to simulate the central carbon pathway in *Escherichia coli* (Chassagnole et al., 2002). The analysis was based on detailed kinetics of enzymes from the phosphotransferase system (PTS, pathway used by bacteria for sugar uptake), glycolysis and pentosephosphate pathway (metabolic pathway parallel to glycolysis that produces NADPH).

In parallel, Metabolic Control Analysis (MCA) was proposed by Kacser and Burns (Kacser and Burns, 1973) and, independently from them by Heinrich and Rapoport (Heinrich and Rapoport, 1974). MCA is a mathematical framework that permits the analysis of the steady-state change in the concentration of a metabolite or in the flux through a metabolic pathway when, for example, the activity of an enzyme from the pathway changes. In this context, steady-state is reached once the concentration of all metabolites in the pathway no longer change. One of the most important findings of MCA is that the control of flux is distributed among all enzymes in a pathway. This finding substituted the intuitive, but incorrect concept of a rate-limiting step. Most often, when a single enzyme from a pathway changes its enzymatic activity, the flux through the pathway changes only modestly (Moreno-Sánchez et al., 2008).

Some 20 years back, whole genome sequencing and gene annotation revolutionized the study of metabolism. At this time, it became relatively easy to find the reactions that comprise a metabolic pathway in an organism of interest. However, the enzymatic analysis required to measure the kinetic parameters of every enzyme from a pathway did not keep up with the speed at which pathways were discovered. Thus, the field shifted from the detailed simulation studies that began in the 70s to an analysis of metabolism based solely on reaction stoichiometry, where the only knowledge needed is the ratios in which substrates are combined and converted to products. In the next section I describe these stoichiometric models in detail.

1.2.2 Stoichiometric modeling of metabolism

Building genome-scale metabolic networks

Today, we can sequence and annotate whole genomes of organisms quite easily. This allows the reconstruction of metabolic networks, which consist of metabolites and of biochemical reactions transforming these metabolites into each other, at genome-scale. This reconstruction can be done even of organisms for which there is little biochemical information. As a result, 20 years after the first genome-scale metabolic network of *Haemophilus influenzae* was published (Edwards and Palsson, 1999), public databases host thousands of metabolic networks for different organisms (5897 bacteria, 127 archaea and 215 eukaryotes (Gu et al., 2019)).

The genome-scale reconstruction of a metabolic network consists of finding a list of all metabolic reactions that take place in an organism. This reconstruction process can be divided into four main steps (Feist et al., 2009). First, a list of metabolic enzymes thought to be present in the organism is inferred from genome annotations. With the help of databases, one determines the metabolic reactions these enzymes carry out. Nowadays, this step of the network reconstruction is done semi-automatically and results in an initial set of candidate metabolic reactions. The second step is more laborious and is essential for a high quality reconstruction. It consists of curating the draft reconstruction. This means establishing organism-specific features such as substrate and enzyme cofactor specificity as well as the sub-cellular localization of enzymes. Third, a *biomass reaction* is established. This reaction drains biomass precursor compounds, such as amino-acids, lipids, nucleotides, and cofactors, in experimentally determined ratios. It is fundamental for the simulation of cell growth with methods such as Flux Balance Analysis (explained below) but it is also essential for testing the quality of the reconstruction. Finally, high-throughput growth data for the organism in different environments and across different genetic conditions is used to identify discrepancies between modeling predictions and experimental data. This information is used to add or remove pathways to improve a metabolic model and increase its predictive power.

Once the reconstruction process is finished one (ideally) has a list of all metabolic reactions taking place in a organism, together with each reaction's stoichiometric coefficients. This information can be mathematically represented as a stoichiometric matrix S with as many rows as metabolites (m) and as many columns as reactions (r). Each entry S_{ij} of the stoichiometric matrix contains the stoichiometric coefficient with which metabolite i participates in reaction j.

Pan-metabolism and random viable networks

To date, the Kyoto Encyclopedia of Genes and Genomes (KEGG) contains 11383 biochemical reactions in its database (Kanehisa et al., 2016) - the universe of known metabolic reactions. The analysis of networks created by taking subsets of reactions from this pan-metabolism may be instructive. For

example, one can create a subnetwork containing only the reactions that are part of a metabolic pathway, to better understand the pathway's functioning (Hädicke and Klamt, 2017; Rawls et al., 2019). Alternatively, one may find broad commonalities of metabolic systems through the analysis of networks that emulate genome-scale networks, but which contain a random complement of reactions that is not observed in any existing organism.

A simple random selection of reactions from the pan-metabolism is unlikely to result in a network capable of synthesizing all chemical building blocks needed for growth. Yet, a method based on Markov Chain Monte Carlo sampling can be efficiently used to find viable metabolic networks with an otherwise random set of reactions (Rodrigues and Wagner, 2009; Samal et al., 2010). Random viable networks have been used to study the evolution of metabolism (Rodrigues and Wagner, 2009) and the role of historical contingency in evolution (Barve et al., 2014). They have also been used to show that evolutionary innovations may occur non-adaptively (Barve and Wagner, 2013), or to study the role of recombination on producing organisms able to survive in new environments (Hosseini et al., 2016).

Pan-metabolisms, its subnetworks and random viable networks can be mathematically represented by stoichiometric matrices just as I described before for the genome-scale metabolic networks of living organisms. The same methods can thus be used to study all these networks. These methods are described in the next sections.

Analysis of metabolic networks based on graph theory

The stoichiometric matrix of a metabolic network can be represented as a graph. A graph is a mathematical structure used to represent connections between pairs of objects. Two objects (or nodes) are connected to each other (by an edge) if the nodes show a specific relationship (Pavlopoulos et al., 2011). For example, a graph could be used to represent airport connectivity if its nodes are airports, and if an edge exists between these nodes if there is a flight between the respective airports.

Hypergraphs are a generalization of graphs (Klamt et al., 2009). In a hypergraph an edge can link any number of nodes. Since most metabolic reactions are multi-molecular, a metabolic network can be viewed as a hypergraph. Nonetheless, metabolic networks are usually represented as graphs. Even though information is lost in this representation, it has the advantage that one can use the extensive tools and algorithms available from graph theory (Bondy and Murty, 1976) which have not been yet developed for the analysis of hypergraphs.

There are various ways in which metabolic networks can be represented as graphs (Takemoto, 2012). One possibility is to use a bipartite graph in which there are two different types of nodes - reactions and metabolites. Edges connect only metabolites to reactions (and vice versa), but these do not connect metabolites with other metabolites and reactions with other reactions. Bipartite graphs can represent all information a hypergraph represents. However, with this type of graph we deal

with the same disadvantage mentioned earlier with respect to hypergraphs, that is, not all available graph-theoretical methods can be used for the analysis of bipartite graphs (Klamt et al., 2009). An alternative is to use substrate or reaction graphs. In the former, nodes correspond to metabolites, and two nodes are connected by an edge if they occur in the same reaction (either as substrate or product). In the latter, nodes correspond to reactions, and two nodes are linked if they share at least one metabolite (either as substrate or product).

Graph-theoretical analysis of metabolic networks has proven informative. For example, it showed that in metabolic networks a few nodes are highly connected while the remaining ones are only more modestly connected (Jeong et al., 2000; Albert, 2005; Broido and Clauset, 2018). It also showed that the distance between any pair of nodes is surprisingly small, a property referred to as the 'small world' property (Watts and Strogatz, 1998; Wagner and Fell, 2001). Also, metabolic networks are composed of small highly connected modules, each carrying discrete functions (Ravasz et al., 2002). In addition, graph-based analysis combined with DNA sequence analysis of metabolic enzyme-coding genes showed that central and highly connected enzymes evolve more slowly than less connected ones (Vitkup et al., 2006).

Constraint-based modeling

Various physicochemical, environmental, and regulatory constraints limit the phenotypic potential of organisms. Constraint-based modeling refers to a group of methods that use these constraints to predict the metabolic phenotype of an organism (Price et al., 2004). Metabolic phenotype is a broad concept used to refer to those observable characteristics of an organism which have to do with metabolism. In the context of metabolic networks, these characteristic refer to, for example, the capacity of an organism to synthesize biomass in certain environment, or the repertoire of metabolites consumed or produced by an organism.

The most important constraint imposed on metabolic networks is that of steady-state, where no metabolite changes its concentration. In other words, the production rate of a metabolite has to be equal to its consumption rate. This steady-state assumption is valid whenever there is time-scale separation between fast intracellular metabolic conversions and slow genetic regulation (Maarleveld et al., 2013). Mathematically, this constraint can be written as $Sv = 0$, where S is the stoichiometric matrix, describing the stoichiometry of the metabolic reactions included in the network; v is a vector whose entries v_i represent the metabolic flux through reaction i in the network. Large networks usually contain more reactions than metabolites; hence there are more unknown variables than equations and there is no unique solution to this system of equations.

Thermodynamic constraints set the directionality of reactions and enzyme capacity constraints limit the maximum rate at which an enzymatic reaction can occur. Both principles cause further

constraints that can be added as lower and upper bounds to the flux through a reaction i: $l_i \leq v_i \leq u_i$. Any other constraint affecting a metabolic network can be added analogously. Constraints are usually not restrictive enough to reduce the flux distribution through a metabolic network to a single possibility. Instead, constraints limit flux distributions to a space of possibilities known as the 'solution space'. In a space with as many dimensions as there are reactions in a network, the solution space corresponds to the volume in that space that fulfills all constraints. In other words, every point in the solution space corresponds to one possible flux distribution compatible with the constraints imposed.

Once all relevant constraints are introduced, it is possible to characterize potential phenotypes that satisfy all constraints. One possibility is to uniformly sample the solution space (Almaas et al., 2004). An alternative is to determine the extreme pathways or elementary modes (Schilling et al., 1999; Papin et al., 2004). These are sets of flux distributions which can be combined to describe all possible solutions in the solution space. These approach have been useful, for example, to study metabolic network redundancy (Price et al., 2002b), robustness of an organism to gene deletion (Stelling et al., 2002), and optimal design of bacterial strains (Carlson et al., 2002).

Alternatively, one can perform an optimization to find, among all potential phenotypes that satisfy all constraints, one that best achieves a certain function. For example, in the widely used method of Flux Balance Analysis (FBA) (Orth et al., 2010), biomass production is usually the function that is maximized to find a flux distribution that results in maximal biomass growth of an organism. The argument behind this optimization strategy is that organisms have been tuned by evolution to achieve maximal growth. Various other optimization strategies, such as maximization of ATP production, or minimization of resource consumption, have also been used (Schuetz et al., 2007).

FBA has been successfully used for predicting biomass growth, nutrient uptake rates, by-product excretion, the effects of adaptive evolution, as well as optimal strain design (Edwards and Palsson, 2000; Edwards et al., 2001; Ibarra et al., 2002; Förster et al., 2003; Varma and Palsson, 1994a; Raman and Chandra, 2009; Gianchandani et al., 2010).

Other commonly used methods to predict metabolic phenotypes based on metabolic constraints and optimization include Flux Variability Analysis (FVA (Gudmundsson and Thiele, 2010)), Minimization of Metabolic Adjustment (MOMA (Segrè et al., 2002)) and Regulatory on/off minimization (RooM (Shlomi et al., 2005)). In FVA the flux through a reaction is both maximized and minimized to find the range of flux values that a reaction can take without violating the constraints imposed. MOMA and RooM were developed to improve predictions of an organism's metabolic phenotype after gene knock-out mutations. Both methods rest on the assumption that it takes time for regulatory changes to occur, and for an organism to reorganize fluxes to grow optimally. They propose that immediately after a gene knock-out mutation occurs, metabolic flux distribution will be sub-optimal and close to that observed in the wild-type. While in MOMA the flux distribution for the knock-out mutant is

found by minimizing the Euclidean distance of the post mutation flux distribution to the wild-type flux distribution, in RooM the number of reactions requiring a flux change is minimized. Lately, a modified version of FBA, named dynamic FBA (dFBA,(Mahadevan et al., 2002)), has gained attention. It couples FBA with ordinary differential equations to quantify microbial growth and resulting changes in environmental composition over time.

1.2.3 Modeling microbial communities

Various modeling strategies have been used for the analysis of microbial communities, each best-suited for a specific problem. Often the dynamics of microbial communities are modeled using differential equations, where typically, a modified version of the well-known Lotka-Volterra model (Lotka, 1910) is used (Stein et al., 2013; Lenski, 1988). The Lotka-Volterra model is a classical ordinary differential equations model. It is used to describe the change of a community of species over time, as a function of their intrinsic growth rates and the interactions between species. When one is interested in the community composition of interacting organisms (which cooperate or compete) which persists unchanged with time, instead of in the whole community dynamics, a game theoretic analysis may be best (Gore et al., 2009; Healey et al., 2016). For this analysis one needs to have information about the pay-off associated with different strategies, such as cooperation or defection. In a game where each "player" (species) can use one of two strategies, each player is awarded a pay-off that depends on the strategy of the other player. A third class of models are individual-based models. They are well-suited, for example, to analyze how spatial structures (such as biofilms) emerge (Nadell et al., 2010; Mitri et al., 2011; Estrela and Brown, 2013; Ghosh et al., 2015). In this modeling framework, each cell is modeled explicitly and the dynamics of the whole community arises from that of individual cells.

Community models based on stoichiometry are often used as a mechanistic model, where interactions between the community members are inferred rather than assumed. For example, communities can be modeled by combining the stoichiometric matrices of individual species into one community-stoichiometric-matrix (Stolyar et al., 2007). The limitation of this analysis lies in the definition of the objective function that a community maximizes. It is common to find conflicts between the objectives of different community members, and the choice of a community-objective is often arbitrary. Other methods have been proposed to overcome this limitation. One of them solves nested FBA problems (Zomorrodi and Maranas, 2012). At first it performs FBA on every community member independently, and then on the community level (usually maximizing the entire community biomass). Another alternative is to use an extension of dynamic FBA, solving FBA independently for each species in the community and updating the environmental composition according to what the species consumed and excreted between rounds of FBA (Zhuang et al., 2011; Hensen and Hanly, 2014).

In the laboratory, experiments are often done in well-mixed environments but microbial communi-

ties may form complex spatial structures in nature. To account for this spatial heterogeneity, partial differential equation models (Datta et al., 2013) as well as extensions of models based on game-theory (Allen et al., 2013) or dynamic-FBA (Harcombe et al., 2014b) are used.

1.3 Thesis outline

Two centuries after the work of Wallace and Darwin, it is still puzzling how all the biodiversity that Earth holds evolved and is maintained. Solving this puzzle, is one of the fundamental aims of biology. In this thesis I contribute to this question by studying the evolution, assembly and persistence of microbial communities. Specifically, I investigate the role that cross-feeding interactions play in these processes.

We study macro-organisms and microbes in fundamentally different ways. We have learned extensively about the evolution and ecology of animals and plants from observing them in their natural habitats. In addition, there is a rich fossil record that permits testing hypotheses about their evolution. In contrast, we cannot directly observe microbes in nature, and their fossil record is sparse. Most of our knowledge on microbes comes from experiments performed in the laboratory, where species are studied in isolation or in small communities. Recently, low costs of sequencing and large amounts of publicly available data have facilitated the study of microbial populations. We now face the problem of how to best sample microbial communities so that we include all organisms with which microbes interact, but exclude those that may be physically close on a macroscopic scale, but that are irrelevant for the coexistence of the species in the community.

From the study of macro-organisms, we have learned that species-species interactions play an essential role in the evolution and maintenance of communities. Experiments such as those described in section 1.1.2, where populations of *E. coli* diversified into stably coexisting strains through evolving cross-feeding interactions, suggest that species-species interactions might be important for the evolution and ecology of microbes as well. Inspired by this observation, I here explore the role of cross-feeding for microbial communities in detail. Identifying species-species interactions such as cross-feeding in microbial communities, either in nature or in the lab is challenging. Fortunately, multiple curated metabolic networks for different organisms are available, together with experimentally validated constraint-based methods that allow us to simulate metabolic systems. In addition, we have broad knowledge on the universe of metabolic reactions, as well as a method to efficiently sample this large universe of reactions to obtain random viable networks. With these resources, metabolic analysis becomes suitable to study microbial communities, because we can infer cross-feeding interactions from the metabolic reactions present in the metabolic network of a species without the difficulties of experimentation.

As stated earlier, the goal of this thesis is to investigate the role that cross-feeding interactions play

in the evolution and ecology of microbial communities. In chapters 2 and 3, I study the function of cross-feeding in microbial evolution. To this end, I explore the repertoire of cross-feeding interactions that could evolve between members of a species, which can facilitate the diversification of a population into coexisting ecotypes that feed on different resources. Moreover, I study the extent of metabolic changes required for the evolution of various cross-feeding interactions. Then, in chapter 4, I switch from the analysis of one species to the analysis of communities. Specifically, I explore how cross-feeding between different species shapes the assembly of communities. In the following paragraphs I describe more extensively the objectives and main results of each chapter.

The aim of the first analysis (presented in **Chapter 2**) is to explore the potential of *E. coli* to diversify in simple environments as cross-feeding interactions evolve between individuals of a population. To this end, I first explore which niches *E. coli* can create when growing in glucose medium as it excretes by-product metabolites. To find out, I use the latest genome-scale metabolic model of *E. coli* together with Flux Balance Analysis. I search for all those carbon-containing metabolites that are by-products of *E. coli* growth in glucose and that could also serve as carbon sources for *E. coli* once consumed. I find 58 such metabolites. Then, I study the stability of two-strain-communities: a strain that consumes glucose but excretes a metabolic by-product into the environment, and a strain that thrives on the by-product excreted by the first strain. For this purpose I use dynamic FBA. Last, with the aim of exploring if the cross-feeding potential that my analysis uncovered is a generality of microbes or a peculiarity of *E. coli*, I repeat my analysis using genome-scale models for two other organisms (*Bacillus subtilis* and *Saccharomyces cerevisiae*), and for hundreds of random viable metabolisms. Every metabolic network analyzed resulted in a repertoire, more or less extensive, of metabolites that could be cross-fed. These results suggest that diversification in sympatry through the evolution of cross-feeding interactions could play an important role in the diversification of microbial populations.

The motivation for the analysis presented in **Chapter 3** comes from the results of **Chapter 2**. Although I had found that 58 cross-feeding interactions could potentially arise in populations of *E. coli* grown in glucose, only two such metabolites have been experimentally observed, those involving acetate and glycerol as metabolic by-products. I hypothesized that the likelihood of a cross-feeding interaction to evolve would be inversely related to the number of mutations required, and that multiple mutations might be required for the evolution of computationally predicted ('the possible') but not experimentally observed ('the actual') cross-feeding interactions. To explore this possibility I developed a method to search for the minimal number of metabolic changes required for individuals to change their metabolic state from an ancestral-like glucose-consuming state to an evolved state that produces or consumes a metabolic by-product. I found that the metabolic changes required for the evolution of acetate and glycerol cross-feeding are not less complex than those required for the evolution of

other computationally predicted cross-feeding interactions. My observations suggest that multiple cross-feeding interactions may await discovery. **Chapter 3** is named **The possible and the actual** in recognition of Nobel laureate François Jacob who in his book of the same name expressed that the speculative and the empirical are not rival, but complementary facets of science. I share Jacob's opinion, which I think is reflected in the third chapter.

In **Chapter 4** I extend my analysis from pairs of cross-feeding strains to communities composed of dozens of species maintained by cross-feeding cascades. In this analysis, the metabolic by-products of one species serve as a resource for other species which in turn excrete by-products that yet other species can profit from. The analysis focuses on the exploration of whether biodiversity inhibits further biodiversity as niches become filled with increasing numbers of species, or whether biodiversity fosters further biodiversity as niches are created through species-species interactions such as cross-feeding. To this end, I perform a mechanistic bottom-up analysis in which species are modeled with random viable networks. Species-species interactions, such as cross-feeding and competition, result from the networks of metabolic reactions I use to model species. I find that communities can host dozens of species in homogeneous environments containing a single initial resource. However, diversity reaches a maximum when no more new niches can be created, and when available niches cannot be partitioned any more finely.

In sum, my work suggests that great microbial biodiversity can arise in sympatry, even in simple environments that offer few resources, as cross-feeding interactions evolve. These cross-feeding interactions may have not only facilitated the evolution of microbes, but may also play essential roles for the persistence of microbial communities.

Chapter 2

Cross-feeding potential in microbes

Published as:

San Roman M. and Wagner A. (2018) An enormous potential for niche construction through bacterial cross-feeding in a homogeneous environment. *Plos Computational Biology*

2.1 Summary

Microorganisms modify their environment by excreting by-products of metabolism, which can create new ecological niches that can help microbial populations diversify. A striking example comes from experimental evolution of genetically identical *Escherichia coli* populations that are grown in a homogeneous environment with the single carbon source glucose. In such experiments, stable communities of genetically diverse cross-feeding *E. coli* cells readily emerge. Some cells that consume the primary carbon source glucose excrete a secondary carbon source, such as acetate, that sustains other community members. Few such cross-feeding polymorphisms are known experimentally, because they are difficult to screen for. We studied the potential of bacterial metabolism to create new ecological niches based on cross-feeding. To do so, we used genome scale models of the metabolism of *E. coli* and metabolisms of similar complexity, to identify unique pairs of primary and secondary carbon sources in these metabolisms. We then combined dynamic flux balance analysis with analytical calculations to identify which pair of carbon sources can sustain a polymorphic cross-feeding community. We identified almost 10000 such pairs of carbon sources, each of them corresponding to a unique ecological niche. Bacterial metabolism shows an immense potential for the construction of new ecological niches through cross feeding.

2.2 Introduction

With as many as one trillion predicted species, microbial diversity on our planet is enormous (Locey and Lennon, 2016). To understand the origins of biological diversity in general and microbial diversity in particular is a central goal of ecology and evolutionary biology. For many decades, most biological diversity was thought to arise in allopatry, that is, when populations become physically subdivided (Coyne, 1992). More recently, biologists have increasingly accepted that populations can also diversify in sympatry, that is, without any physical barriers (Kondrashov and Kondrashov, 1999; Dieckmann and Doebeli, 1999a; Higashi et al., 1999; McKinnon and Rundle, 2002; Barluenga et al., 2006; Meyer et al., 2016; Feder et al., 1988). Examples of sympatric diversification include insect populations that adapt evolutionarily to different plant hosts (Feder et al., 1988), stickleback populations that evolve reproductive isolation at least partly in sympatry (McKinnon and Rundle, 2002), Midas cichlid populations that originated in a small volcanic crater lake in Nicaragua (Barluenga et al., 2006), and bacteriophage lambda that specializes on different bacterial hosts (Meyer et al., 2016). In bacteria, sympatric divergence has been observed both in nature (Sikorski and Nevo, 2005; Coleman et al., 2006) and during experimental evolution (Rainey and Travisano, 1998; Helling et al., 1987; Rozen and Lenski, 2000a; Koeppel et al., 2013).

Sympatric diversification is easiest in heterogeneous environments (Smith, 1966; Kassen, 2002).

Because such environments provide multiple ecological niches, organisms can easily diversify when they specialize and adapt to these niches. Such diversity can then be maintained according to the niche exclusion principle – the principle states that different organisms cannot occupy the same niche (Gause, 1934). Examples include the spatial structure of an unshaken growth medium, which facilitates morphological diversification in experimental evolution of *Pseudomonas fluorescens* (Rainey and Travisano, 1998); spatial (free-living or particle-associated) and temporal (spring and fall) resource partitioning, which triggers sympatric speciation in bacterioplankton (Hunt et al., 2008); the divergence that occurs as a result of host shifts from hawthorn to domestic apples in apple maggot flies (Feder et al., 1988); as well as the specialization of bacteriophages to *Escherichia coli* expressing different membrane proteins (Barluenga et al., 2006; Feder et al., 1988).

In apparent contradiction to the niche exclusion principle, sympatric diversification can also occur in homogeneous environments (McKinnon and Rundle, 2002; Barluenga et al., 2006; Sikorski and Nevo, 2005; Coleman et al., 2006; Helling et al., 1987; Rozen and Lenski, 2000a; Good et al., 2017). Perhaps the most striking example involves stable genetic polymorphisms that can originate in *E. coli* populations cultured in the homogeneous and well-mixed environment of a batch culture or a chemostat, a device in which a cell culture is kept in a constant nutrient environment by continually supplying it with nutrient medium (Helling et al., 1987; Rozen and Lenski, 2000a; Good et al., 2017; Kinnersley et al., 2014; Rosenzweig et al., 1994; Treves et al., 1998). For example, over a mere 800 generations of laboratory evolution in a glucose-limited chemostat, initially isogenic populations of *E. coli* can diversify into multiple genetically different strains (Helling et al., 1987; Kinnersley et al., 2014; Rosenzweig et al., 1994; Treves et al., 1998). These strains stably coexisted in the chemostat as a result of cross-feeding (Rosenzweig et al., 1994). That is, one strain consumes the primary carbon source glucose and excretes a secondary carbon source (acetate or glycerol), whereas the other strain feeds on the secondary carbon source. These phenotypic differences result from regulatory DNA mutations in transcription factors and cis-regulatory regions. They include a cis-regulatory mutation affecting the expression of acetyl CoA synthetase, an enzyme that catalyzes the transformation of acetate to acetyl CoA, which enters the tricarboxylic acid cycle to produce energy. They also include a structural mutation in the glycerol-3-phosphate repressor, which can result in constitutive expression of glycerol utilization genes (Kinnersley et al., 2014). Experiments like this suggest that *E. coli* may readily diversify genetically and metabolically in a completely homogeneous environment.

The emergence of cross-feeding is an example of niche construction, a process where organisms change their environment in ways that can affect the evolutionary dynamics of themselves and of other organisms (Laland et al., 1999; Day et al., 2003; Erwin, 2008; Thakur and Wright, 2017; Matthews et al., 2014; Laland et al., 2016). Prominent examples of niche construction include animals that construct artifacts such as webs, nests and burrows (Laland et al., 1999); earthworms and plants that

alter the fertility, humidity and chemical composition of soil (Matthews et al., 2014; Turner, 2009; Holmgren, M. et al., 1997); and bacteria that construct biofilms and excrete antibiotics as well as metabolic by-products (McNally and Brown, 2015). Constructed niches can affect evolution even on the short time scales of experimental evolution, where populations of *Pseudomonas fluorescens* become dependent on their own modifications of their chemical environment (Callahan et al., 2014; Loudon et al., 2016).

The origin of new niches associated with bacterial cross-feeding is not easy to detect experimentally: Except for differences in colony morphology, cross-feeding polymorphisms generally lack phenotypes that are both macroscopically visible and highly specific. However, computational analysis can help predict the conditions under which cross-feeding polymorphisms can originate and persist. Some authors use small biochemical networks to search for the conditions that would promote genetic diversification through cross-feeding interactions (Doebeli, 2002; Pfeiffer and Bonhoeffer, 2004; Gudelj et al., 2016, 2007; Porcher et al., 2001; Keymer et al., 2012). Others use digital organisms with evolvable genomes and metabolic networks (Rocabert et al., 2017). Yet others simulate individuals in an evolving population where random mutations can change nutrient consumption rates in a model of *E. coli* central carbon metabolism, and show that glucose-acetate cross-feeding can originate in such a population. Most recently, a genome scale metabolic network of *E. coli* was used to study cross-feeding and other metabolic dependencies that emerge as a result of evolution under gene loss (McNally and Borenstein, 2017) or amino-acid leakage (Zomorrodi and Segrè, 2017).

Here we go beyond this work and evaluate the general potential for the construction of new niches associated with cross-feeding that is inherent to the metabolism of *E. coli* and to complex metabolic systems in general. That is, we ask how many different kinds of ecologically stable cross-feeding interactions can emerge in an initially homogeneous population, where one bacterial strain feeds on a primary carbon source and produces a secondary carbon source that sustains the other strain. To answer this question, we take advantage of a well-studied and experimentally validated (Orth et al., 2011) genome-scale model of *E. coli* metabolism. We use Flux Balance Analysis (FBA), an experimentally validated computational technique (Orth et al., 2010), to characterize the production of secondary carbon sources that can help cross-feeding polymorphisms emerge. We then use dynamic flux balance analysis (dFBA) (Varma and Palsson, 1994b; Chiu et al., 2014), a variant of FBA that uses genome-scale metabolic information to predict the ecological dynamics of microbial communities and how they change their chemical environment over time. We use dFBA to study the conditions under which two cross-feeding strains can establish a stable community in a chemostat. After having reproduced the experimentally observed glucose-acetate cross-feeding polymorphism (Helling et al., 1987), we then identify additional pairs of primary and secondary carbon sources that can lead to the establishment of stable cross-feeding communities. We find thousands of such pairs, both in *E.*

coli and other metabolic reaction networks of similar complexity. Our work demonstrates the great potential of metabolic systems to construct new ecological niches.

2.3 Results

2.3.1 The model

Our first analysis prepares the ground by examining the conditions under which a glucose-acetate cross-feeding polymorphism can be stably maintained by two *E. coli* strains. We studied this specific polymorphism, because it is experimentally well documented (Helling et al., 1987; Kinnersley et al., 2014; Rosenzweig et al., 1994; Treves et al., 1998), and aimed to reproduce it in silico. Specifically, we simulated the dynamic of a community composed of two cross-feeding *E. coli* strains (or ecotypes (Koeppel et al., 2008)), a producer strain P that produces a secondary carbon source as a by-product of feeding on some primary carbon source, and a consumer strain C that consumes this secondary carbon source. We use the same genome-scale metabolic network of *E. coli* iJO1366 (Orth et al., 2011) to model both strains. This means that the metabolic networks of both strains comprise exactly the same reactions and metabolites. This modeling decision reflects the observation that cross-feeding strains can emerge from a single *E. coli* ancestor in little evolutionary time (Rosenzweig et al., 1994). The metabolic differences between cross-feeding strains do not result from differences in their complement of enzyme-coding genes, but from regulatory mutations that affect how much of a specific carbon source each strain can consume or produce (Kinnersley et al., 2009).

We model these differences phenomenologically, through differences in the flux through two specific reactions in strains P and C. Specifically, we model the secondary carbon source production of strain P by imposing a non-zero production flux $p_{scs,P}$ for this carbon source via the exchange reaction that transports the secondary carbon source out of the cell. And we model the secondary carbon source consumption of strain C by limiting the strain's primary carbon source consumption. This modeling decision is motivated by the experimental observation that when cross-feeding emerges in *E. coli* (Rosenzweig et al., 1994), the consumer strain's ability to consume its primary carbon source becomes impaired. One might argue that increasing the consumption of the secondary carbon source might be biologically more sensible. However, the two approaches are equivalent. Here is why. Since we simulate a chemostat culture, once steady state is reached, the strains in the chemostat grow at a constant rate. The consumer strain C achieves this growth rate by consuming both primary and secondary carbon sources. If consumption of the primary carbon source increases, consumption of the secondary carbon source becomes reduced by an equivalent amount (Equation 2.8, Supplementary 2.6.4 Text), such that the steady state is unaffected.

For the well-studied cross-feeding interaction of acetate producer and consumer strains, regulatory

mutations in specific genes are known to bring forth the metabolic behavior of producer and consumer strains (Kinnersley et al., 2014; Treves et al., 1998; Plucain et al., 2014). Since the objective of our work was to study not only glucose-acetate cross-feeding but multiple other cross-feeding interactions we decided not to incorporate assumptions about specific mutations in specific genes into our model. By imposing general constraints on the production and consumption of specific carbon sources, we allowed for the possibility that our modeled strains could achieve these constraints in different ways, depending on the carbon source considered. The specific mutations that may underlie our strains' metabolic behavior will be the subject of future work.

2.3.2 Acetate production creates a two-dimensional ecological niche that can stably support two *E. coli* strains through cross-feeding

In the first part of our analysis, we focus on glucose as a primary carbon source, and on acetate as a secondary carbon source. The secondary carbon source is excreted by the producer strain P at a rate $p_{ac,P}$, and consumed by the consumer strain C (Figure 2.1A). We initially assume that the consumer strain C cannot consume glucose ($c_{glc,C} = 0$), an assumption that we relax below (Supplementary 2.6.3 Text). To find out whether both strains can coexist in a stable chemostat community, we first use Flux Balance Analysis (Orth et al., 2010) (FBA, Methods) in the form of dynamic FBA (Varma and Palsson, 1994b; Chiu et al., 2014) (Methods).

To mimic typical experimental conditions, we performed all simulations with a dilution rate D, the rate at which culture is replaced with fresh medium, of D=0.2 h^{-1} (Helling et al., 1987). At this dilution rate, the maximum rate at which *E. coli* cells can produce acetate (p_{ac}^{max}) without being eventually flushed out from the chemostat is 50.3 $mmol\, gDW^{-1}\, h^{-1}$ (Supplementary 2.6.2 Text). (Here and below, all units of metabolic flux are given in $mmol\, gDW^{-1}\, h^{-1}$). To ensure survival of the producer strain P, we simulated chemostat dynamics at an acetate production rate $p_{ac,P}$ that is equal to 5% of this maximum (2.6 $mmol\, gDW^{-1}\, h^{-1}$). We initialized the chemostat in the presence of only the acetate producing strain P, and once this strain had reached steady-state, which occurred after no more than 50 hours, we introduced the acetate consuming strain C. We then monitored the joint dynamics of both strains until they had reached steady-state or until one strain had gone extinct.

Figure 2.1B shows the change in biomass of P and C over time. Only three carbon-containing metabolites – glucose, acetate, and carbon dioxide – change their concentration (Figure 2.1C).The concentration of glucose (Figure 2.1C, grey) decreases as the acetate producer P consumes glucose and this decrease is concurrent with an increase in P's biomass (Figure 2.1B, blue). Strain P metabolizes glucose partially to carbon dioxide (Figure 2.1C, green) and partially to acetate (Figure 2.1C, black), which is why the concentration of both metabolic by-products increases. Once the acetate consumer strain C is introduced into the chemostat at 50 hours (Figure 2.1B, red), the chemical environment

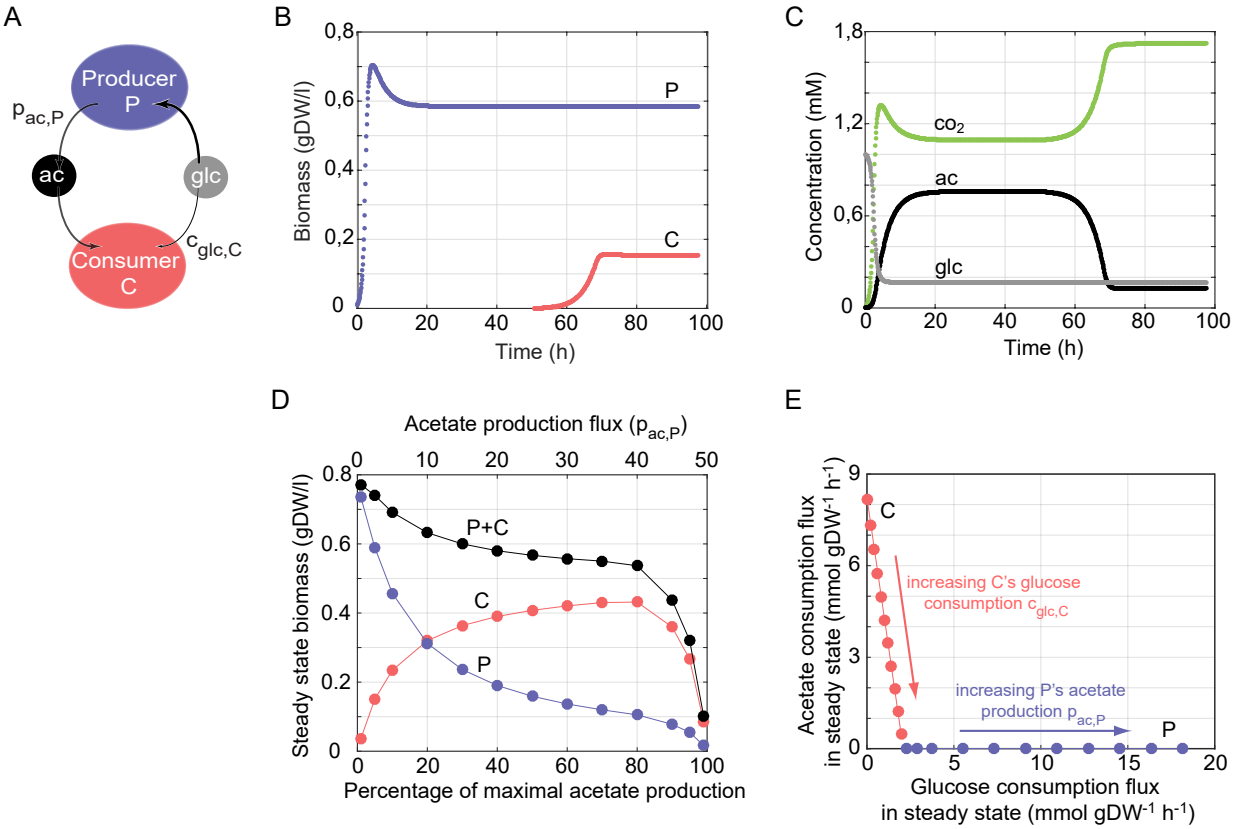


Figure 2.1: Ecological dynamics of an acetate producer *E. coli* strain P and an acetate consumer strain C in a chemostat. (A) Interactions between the strains. Producer strain P (blue) produces acetate (black) at a rate $p_{ac,P}$ and consumes glucose (grey) as its sole carbon source. Consumer strain C consumes mainly acetate but can also consume glucose at some rate $c_{glc,C}$. Both strains may also consume other nutrients or produce other metabolic by-products, which are not shown. (B) Dynamics in a chemostat for an acetate production flux by P of 2.6 ($p_{ac,P}$=2.6 $mmol\,gDW^{-1}\,h^{-1}$). The horizontal axis shows time, and the vertical axis shows the biomass of P (in blue) and C (in red) vs. time. (C), as in (B), but the vertical axis shows the concentration of glucose (grey), acetate (black) and carbon dioxide (green). (D) Steady state biomass (vertical axis) of the producer strain (P, in blue), consumer strain (C, in red), and both strains (P+C, in black) as a function of the acetate production rate (horizontal axis). This rate is expressed either in absolute flux units (top horizontal axis) or as the percentage of the maximal acetate production rate (bottom horizontal axis), that is, the rate beyond which the producer strain grows so slowly that it is flushed out of the chemostat. (B), (C) and (D) show the results of simulations of ecological dynamics in a chemostat inhabited by an acetate producer strain P and an acetate consumer strain C. For the purpose of these figures, it is assumed that strain C cannot consume glucose ($c_{glc,C}$=0). (E) Nutritional niche of the producer strain P (blue) and the consumer strain C (red) when metabolically distinguishable strains coexist. The horizontal and vertical axes show the glucose and acetate consumption rates, respectively, of the indicated strains in metabolic steady-state. P's steady-state nutrient consumption increases with its acetate production $p_{ac,P}$ (blue arrow). The blend and ratio of nutrients that C consumes varies with the maximal glucose consumption rate $c_{glc,C}$. That is, increasing C's maximal glucose consumption $c_{glc,C}$, increases C's glucose consumption in steady state while reducing its acetate consumption, as indicated by the red arrow.

contains a substantial amount of acetate, which strain C metabolizes to carbon dioxide to synthesize biomass. By 100 hours, the system has reached a new steady state, in which a stable polymorphism of the acetate producer (P) and the acetate consumer (C) strain is maintained as a result of their cross-feeding interaction.

We next wanted to find out how the population's behavior changes if the amount of acetate excreted by the producer strain P varies. We thus varied the acetate production rate $p_{ac,P}$ up to the maximum beyond which the producer goes extinct. Not surprisingly, the steady-state biomass of strain P is reduced as its acetate production increases (Figure 2.1D, blue), because of the metabolic cost incurred by acetate production. In contrast, the steady-state biomass of the consumer strain C has a unimodal distribution, with a maximum biomass reached at approximately 80% of the maximal acetate production rate. The reason is that C's biomass reflects the acetate concentration in the chemostat, and this concentration depends not only on the amount of acetate produced per unit of producer strain ($p_{ac,P}$), but also on the amount of producer biomass. As acetate production $p_{ac,P}$ increases, the amount of acetate produced per unit biomass increases but the amount of producer biomass decreases. The joint effect of these opposing patterns is a unimodal distribution of consumer biomass.

The total (community's) biomass (Figure 2.1D, black) decreases with increasing acetate production and has its maximum (0.78 gDW/l) in the absence of acetate production. The reason is that part of the acetate excreted into the chemostat environment is removed through the dilution flux D and not available for usage. In addition, even if all produced acetate were available, its production and later consumption are associated with losses in terms of energy and carbon atoms.

Regardless of the acetate production of the producer strain P, the producer and consumer strains stably coexist, as has also been found experimentally (Rosenzweig et al., 1994). Additional simulations show that the eventual steady-state composition of the chemostat does not depend on the initial biomass of either strain or the time at which C is introduced (Supplementary 2.4 Fig). In contrast, a higher dilution flux D will result in higher biomass for the producer P, but a lower biomass for the consumer C. This can be intuitively understood if we consider the concentration of the nutrients that support growth of each strain: The higher the dilution rate is, the more similar the composition of the chemostat is to that of the fresh medium, which contains high amounts of glucose but no acetate.

When the consumer strain C can also metabolize the primary carbon source ($c_{glc,C} > 0$), the two strains compete for this carbon source, and coexistence is no longer guaranteed. However, analytical calculations supplemented by simulations show that the two strains can stably coexist under a broad range of glucose consumption and acetate production rates (Supplementary 2.6.3 and 2.6.4 Texts, supplementary 2.5 and 2.6 Figs). When they do, they occupy distinct ecological niches in their nutrient environment (Raubenheimer and Simpson, 1999; Behmer and Joern, 2008), as shown in Figure 2.1E. One can visualize the ecological niche space in our chemostat environment as a multidimensional space,

where each axis of the space corresponds to the availability or consumption of a nutrient available in the environment. Because in our analysis only two carbon sources are present, the producer strain P and the consumer strain C can compete only for these two carbon sources, which renders our niche space two-dimensional (Figure 2.1E). Its axes correspond to glucose and acetate consumption rates in metabolic steady state. The ecological niches for our two strains can overlap at the level of glucose consumption (Figure 2.1E and supplementary 2.6 Fig), but the strains cannot consume identical amounts of glucose without losing their metabolic differences. Thus, their ecological niches cannot overlap completely, consistent with the competitive exclusion principle from ecological theory (Gause, 1934). We note that this conception of a niche is consistent with the geometric framework of nutritional niche representations (Raubenheimer and Simpson, 1999; Behmer and Joern, 2008), where niches correspond to the "blend and ratio of nutrients that maximize fitness".

2.3.3 Growth on glucose can create multiple additional carbon-source niches

Our analysis so far reproduced the experimentally observed construction of the glucose-acetate niche (Rozen and Lenski, 2000a; Rosenzweig et al., 1994), and identified the conditions under which two *E. coli* strains can coexist in this niche (Figure 2.1D, supplementary 2.6.3 and 2.6.4 Texts, 2.5 and 2.6 Figs). We next turn to secondary carbon sources other than acetate. Although only glycerol has been experimentally identified as an additional secondary carbon source in cross-feeding experiments (Rozen and Lenski, 2000a; Rosenzweig et al., 1994), *E. coli* cells can produce many other metabolites when growing on glucose (Paczia et al., 2012). These metabolites, as well as possibly additional, still unknown metabolites, might serve as secondary carbon sources. To identify all possible secondary carbon sources, we first identified all carbon containing metabolites in the iJO1366 metabolic network that can be transported across the cell wall (i.e., metabolites containing an associated exchange reaction). We used FBA to identify which of these molecules can sustain *E. coli* growth when present as the sole carbon source, which is a prerequisite for a molecule's usefulness in the cross-feeding interactions we study. FBA predicts 180 metabolites (whose acronyms are given in the circle of Figure 2.2A) that can sustain growth of *E. coli* when used as sole carbon sources (Methods). (See Supplementary Table 3 in (Orth et al., 2011) for a list of standard metabolite acronyms used in Figure 2.2A).

For each of these 180 metabolites, we determined whether *E. coli* can produce the metabolite when it is provided with glucose as the sole carbon source (See methods). Figure 2.2A shows a graphical representation of the answer, where an arc connects glucose (grey arrow) to another carbon source if that carbon source can be produced when glucose is the sole primary carbon source. (This means that the enzymes needed to transform glucose into the carbon source are present in *E. coli*). There are 58 such secondary carbon sources, acetate (black arrow) being one of them. In other words, *E. coli* cells growing on glucose can modify the environment by producing 58 alternative nutrients, each of which

can sustain the life of other *E. coli* individuals.

The secondary carbon sources differ greatly in the maximum rate (p_m^{max}) at which they can be produced (Figure 2.2B), which ranges from 4.75 $mmol\, gDW^{-1}\, h^{-1}$ for N-Acetyl-D-glucosamine(anhydrouds)N-Acetylmuramic acid to 178 $mmol\, gDW^{-1}\, h^{-1}$ for formate. Acetate's maximal synthesis rate is 50.3 $mmol\, gDW^{-1}\, h^{-1}$, about twice the mean production rate of 24.7 $mmol\, gDW^{-1}\, h^{-1}$, and second highest among all secondary carbon sources (together with glycolate). This maximum production rate reflects the cost of producing a carbon source: The costlier the production of a secondary carbon source is, the smaller is its maximum production rate (supplementary 2.7 Fig). In addition, the secondary carbon sources differ in their specific biomass yield α, which is the growth rate that can be achieved per unit of carbon source consumed (see Figure 2.2B and Supplementary 2.6.2 Text and supplementary 2.10 Fig green dots). This yield varies from 0.0014 to 0.30 (in $gDW\, mmol^{-1}$ of carbon source). Acetate's biomass yield equals 0.025 and is thus low, less than half of the mean value of 0.068. Thus, while acetate is not costly to produce, its low biomass yield also does not allow for much biomass production in strains that consume it.

Following our analysis of the glucose-acetate niche (Figure 2.1), we then asked whether glucose, in combination with each of these individual metabolites, could lead to a stable cross-feeding polymorphism. In other words, are there values of the production rate of metabolite x ($p_{x,P}$) and the consumption rate of glucose ($c_{glc,C}$) that lead to a stable cross-feeding polymorphism? We found that all secondary carbon sources other than formate can sustain a stable community of two strains, even though these communities vary greatly in the amount of total biomass that they contain (Figure 2.2C).

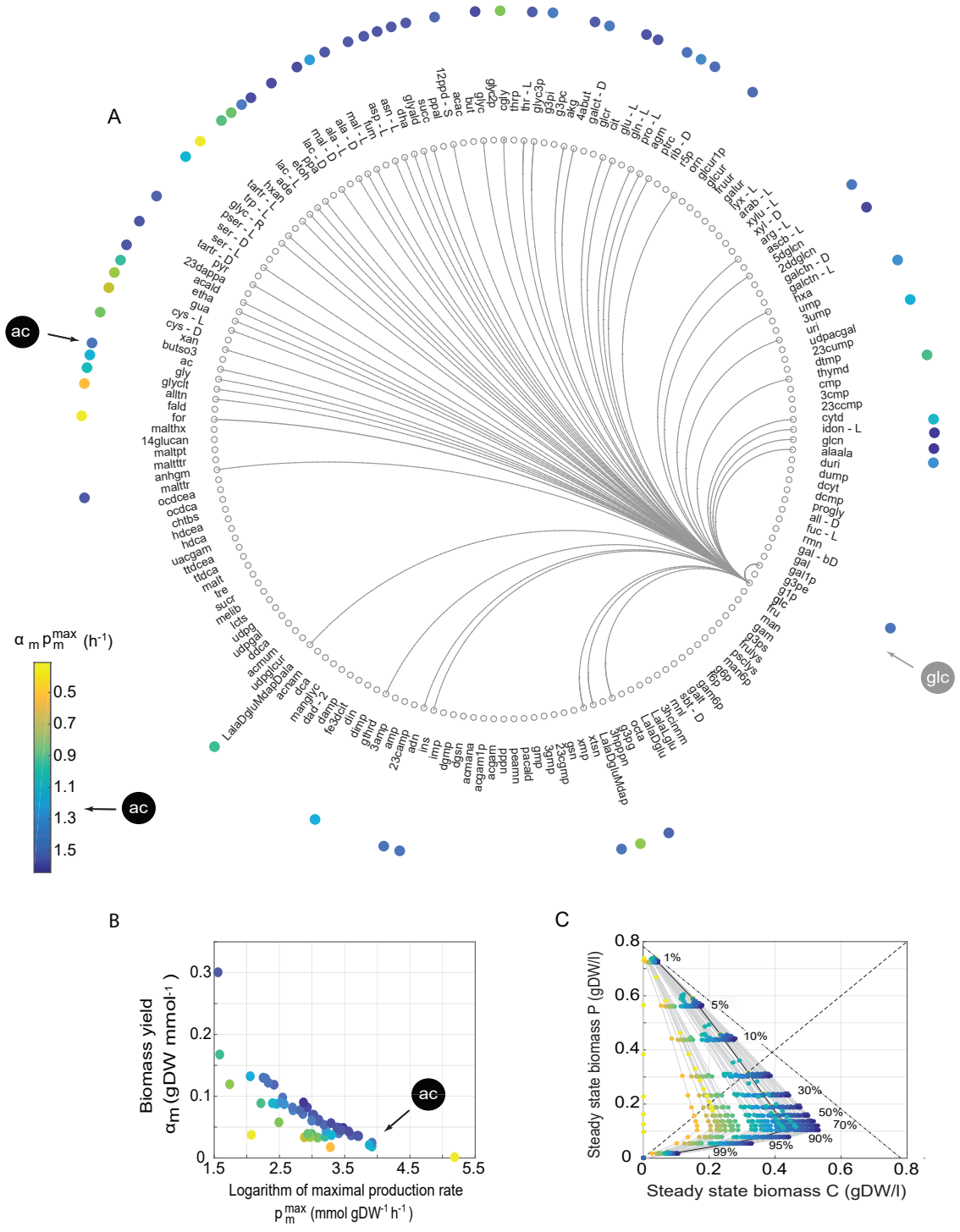

A
ac
glc
α m p max m (h-1)
0.5
0.7
0.9
1.1
1.3
1.5
ac
B
Biomass yield
αm (gDW mmol-1)
0.3
0.2
0.1
0
1.5
2.5
3.5
4.5
5.5
ac
Logarithm of maximal production rate
p max m (mmol gDW-1 h-1)
C
Steady state biomass P (gDW/l)
0.8
0.6
0.4
0.2
0
1%
5%
10%
30%
50%
70%
90%
95%
99%
0
0.2
0.4
0.6
0.8
Steady state biomass C (gDW/l)

Figure 2.2 ***(previous page)*****: Multiple possible cross-feeding interactions involving glucose as a primary carbon source.** (A) Each of the 180 small grey circles labeled with an acronym corresponds to a carbon source that can sustain viability of *E. coli* (iJO1366) when present as a sole carbon source. Metabolites are ranked by increasing biomass yield, starting with formate at 9' o'clock. Arcs connect glucose (glc, grey arrow) with a carbon source (circles) m, if m can be produced when *E. coli* grows on glucose as the sole carbon source. The black arrow indicates the location of the secondary carbon source acetate (ac). Colored circles near each secondary carbon source represent the product of maximal production (p_m^{max}) and biomass yield (α_m) of the carbon source. The magnitude of this product is represented by color (color bar), and this color encoding is applied to all three panels of the figure. (B) Biomass yield and maximal production flux for each secondary carbon source on glucose. The black arrow indicates the circle corresponding to acetate. (C) Steady state biomass values of producer strain P (vertical axis) and consumer strain C (horizontal axis) at different percentages of the maximum synthesis rate (p_m^{max}), i.e., the synthesis rate of the secondary carbon source beyond which the producer strain P is flushed out of the chemostat, for all secondary carbon sources (grey lines). Circles are placed at 1, 5, 10-90, 95 and 99% of the maximum synthesis rate p_m^{max}, as indicated by the numbers in the panel. The dashed-dotted line indicates the total steady-state biomass that is maximally achievable (0.78 gDW/l, obtained when glucose is metabolized completely to CO2 without synthesis of any secondary carbon source). The dashed line indicates where both strains have identical biomass. The product of maximal production (p_m^{max}) and biomass yield (α_m) equals 1.26 h^{-1} for acetate (black line superimposed with blue circles).

If strain P completely respires glucose to carbon dioxide and thus does not excrete any secondary carbon source, the total community biomass is equal to the biomass of P, and reaches a maximum value of 0.78 gDW/l, which is indicated by a dashed-dotted line in Figure 2.2C. For any one secondary carbon source excreted by strain P, the steady-state biomass of the community will change with the carbon source's excretion rate, up to a sustainable maximum (p_m^{max}) beyond which the producer grows so slowly that it will eventually be flushed out from the chemostat. The figure indicates this change by one grey line (and superimposed colored circles) for each of the 54 secondary carbon sources, along with percentages that indicate the percentage of the maximum rate p_m^{max} at which strain P produces the secondary carbon source. (Displaying secondary carbon source excretion as a percentage of this allowable maximum has the advantage of displaying the same cost for all producers, regardless of which secondary carbon source they produce, such that at any given percentage of this maximum, producers of all secondary carbon sources reach the same steady-state biomass.) As the amount of a secondary carbon source produced by P increases, the total community biomass decreases, i.e., the circles in Figure 2.2C become further removed from the dashed-dotted line. We already observed this behavior for acetate (Figure 2.1C, black line in Figure 2.2C), but Figure 2.2C illustrates that it holds for all secondary carbon sources.

The absence of a stable community in the glucose-formate niche space is a consequence of formate's low biomass yield, which requires high formate consumption (112 $mmol\, gDW^{-1}\, h^{-1}$) to support a

growth rate greater than the dilution rate D=0.2 h^{-1}. This level of consumption is impossible under our assumed transport limit (V_{max}= 20 $mmol\, gDW^{-1}\, h^{-1}$). Higher transport limits or lower dilution rates would, however, permit the existence of a stable community.

The steady-state biomass changes of both strains P and C with increased production of a secondary carbon source (Figure 2.2C) are analogous to what we observed in Figure 2.1D, and they exist for the same reason. To support higher production fluxes of any secondary carbon source, P needs to consume more glucose and therefore reaches lower steady-state biomass. To understand the change in steady-state biomass of strain C with an increasing production rate of the secondary carbon source, one has to take into account two factors. The first is the maximal production rate of the secondary carbon source (p^{max}), which affects the carbon source's availability for C's consumption. The second is the biomass yield α of the secondary carbon source, which affects the growth rate achieved per unit flux of consumed carbon source. The product of production and yield ($\alpha_m p_m^{max}$) determines the steady-state biomass of C. If a metabolite is costly (with low maximal production) one can expect low excretion, but a high biomass yield of the same metabolite may compensate for its low production and permit a higher steady-state biomass of C. The colors in Figure 2.2 indicate the magnitude of $\alpha_m p_m^{max}$ for all 58 secondary carbon sources. The figure illustrates that secondary carbon sources whose product of production and yield ($\alpha_m p_m^{max}$) is low will lead to communities with lower total biomass.

The product of production and yield $\alpha_{ac} p_{ac}^{max}$ for acetate does not have an unusually large value (equals 1.26 h^{-1}). About half of the secondary carbon sources (29 out of the 58) have a higher maximal growth rate $\alpha_m p_m^{max}$ than acetate ($\alpha_m p_m^{max} >$ 1.26, blue dots in Figure 2.2D), and therefore support higher community biomass.

The observations in Figure 2.2D are based on the assumption that strain C consumes only the secondary carbon source, but not the primary carbon source glucose ($c_{glc,C}$=0). However, relaxing this assumption to $c_{glc,C}>0$ also allows for stable coexistence of P and C (supplementary 2.9 Fig). The key difference is that stable coexistence then becomes possible for each of the 58 secondary carbon sources, including formate. If C's glucose consumption is so high that it covers at least 90% of the energy and carbon required for persistence in the chemostat, formate can supply the remaining energy and carbon needed. In this case, coexistence of a formate producer strain and a glucose-formate consumer strain would be possible.

2.3.4 Primary carbon sources different from glucose can help construct even more novel niches

Because glucose is not the only primary carbon source that can sustain *E. coli*, we extended the previous analysis of searching for secondary carbon sources from glucose to all 180 primary carbon sources. We began by identifying the number of potential secondary carbon sources that can be produced when

	E. coli	*B. subtilis*	*S. cerevisiae*	Pan-metabolic network
Reactions	2583	1250	1577	7222
Metabolites	1805	990	1226	5625
Exchange reactions	330	229	164	330
Primary carbon sources	180	119	52	221
Secondary carbon sources (on glucose)	58	35	31	86
Secondary carbon sources (total)	83	46	34	109
Pairs of primary-secondary carbon sources	9913	4146	1585	18959
Blocked reactions	227	291	553	3070

Table 2.1: Metabolic characteristics of *E. coli*, *B. subtilis*, *S. cerevisiae* and the pan-metabolic network.

E. coli grows on each primary carbon source. This number ranges from 54 to 62, depending on the primary carbon source (blue circles in supplementary 2.10 and 2.11B Figs).

Most secondary carbon sources can be produced from all primary carbon sources, as is the case for acetate, but some secondary sources can be produced from just a few primary carbon sources (red circles in supplementary 2.10 and 2.11B Figs).

Our analytical results (supplementary 2.6.4 Text) reveal that coexistence is possible for each pair of primary and secondary carbon sources, as long as two conditions are met. The producer strain must produce the secondary carbon source, and the consumer strain C must be able to persist in the chemostat by consuming both the primary and the secondary carbon source (not just the primary carbon source alone). If this were not the case, that is, if the consumer strain was able to persist in the chemostat by consuming only the primary carbon source, then it would have an advantage over the producer strain, which uses part of the consumed primary carbon source to produce the secondary carbon source. In this case, the producer strain would go extinct.

In total, our analysis finds 83 different secondary carbon sources and 9913 unique pairs of primary and secondary carbon sources that allow stable coexistence of a producer strain P and a consumer strain C (Table 2.1). Taken together, these observations imply that the synthesis of by-product metabolites by *E. coli* can create an enormous number of new ecological niches whose identity depends on the primary carbon source available in the environment.

2.3.5 The potential for metabolic niche construction is not a peculiarity of *E. coli* metabolism

The metabolism of any one organism is the product of a long evolutionary history. *E. coli*'s enormous potential for the creation of novel metabolic niches could be an accident of this evolutionary history, or it could be a more general property of the chemical reaction networks that constitute a metabolism. To

find out, we performed several additional analyses. First, we analyzed the niche construction potential of two microbes different from and not closely related to *E. coli*, i.e., the soil bacterium Bacillus subtilis (model iYO844 (You-Kwan Oh et al., 2007)) and the yeast Saccharomyces cerevisiae (model iMM904, (Mo et al., 2009)). (See methods for a detailed description of the procedure.) The analysis revealed (Table 2.3) that these organisms also have a large niche construction potential. They can form stable cross-feeding communities with more than 1000 pairs of primary and secondary carbon sources (Table 2.1).

Each of these three organisms has its own evolutionary history which molded its metabolic network. The observation that they all share a large potential to construct new metabolic niches hints that this potential is a general property of metabolic systems, and not just a peculiarity of the organisms studied and their evolutionary history. To exclude the influence of this history more rigorously, we repeated our analysis with metabolic networks that are not the product of evolution, but that we created in silico with an algorithm that produces random viable networks. These are biochemical reaction networks that produce all essential biomass molecules in a given chemical environment, but contain an otherwise random complement of biochemical reactions drawn from the known "universe" of such reactions. We obtained these networks through a previously published (Samal and Martin, 2011; Rodrigues and Wagner, 2009) Markov Chain Monte Carlo (MCMC) procedure that samples such networks from a vast space of metabolic networks (See Methods).

We note that simpler sampling methods, such as "brute force" uniform sampling of a given number of reactions from a reaction universe is very unlikely to yield viable networks (Samal et al., 2010). In contrast, MCMC sampling can yield not only viable networks but viable networks whose reaction complement is effectively random beyond the requirements imposed by viability, as shown by previous work (Samal et al., 2010; Samal and Martin, 2011).

We used this method to create samples of 500 random viable networks viable on glucose as a sole carbon source and that have the same number of reactions (2251) as the *E. coli* network. We made these networks permeable to all 330 metabolites to which *E. coli* is permeable. In other words, these random viable metabolisms have the potential to consume and produce the same metabolites as *E. coli*.

In our sampling procedure, we only required these networks to be viable on glucose, but as a result of complex correlations between metabolic phenotypes (Barve and Wagner, 2013; Hosseini and Wagner, 2016), they are usually also viable on additional primary carbon sources (Figure 2.3A). Specifically, the number of primary carbon sources on which each sampled metabolic network is viable ranges from 1 to 52 (mean 32±10) (Figure 2.3A). We observe 218 primary carbon sources on which at least one of these random networks is viable.

When exposed to glucose as the primary carbon source, these networks produce between 0 and 22

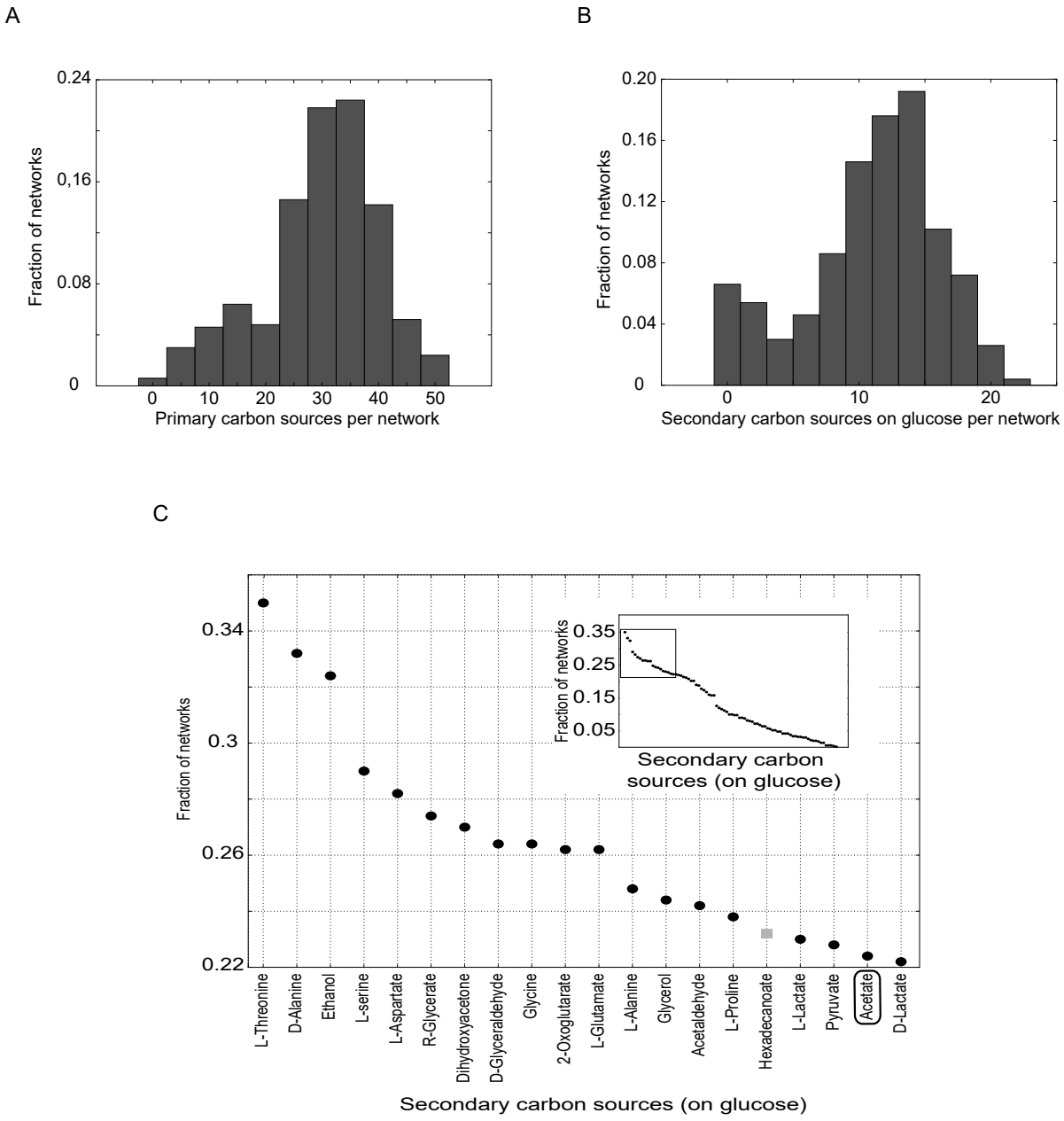


Figure 2.3: Niche construction potential with generic metabolic systems. (A) Histogram of the number of primary carbon sources on which random metabolic networks required to be viable on glucose are also viable. (B) Histogram of the number of secondary carbon sources produced by random metabolic networks required to be viable on glucose. (C) Rank plot of secondary carbon sources produced by at least one random viable network when glucose is used as a primary carbon source, ranked by the fraction of random viable networks by which the secondary carbon source is synthesized. The main panel shows the names of the 20 metabolites with the highest rank, which includes acetate (black box) and glycerol. The grey square shows one secondary carbon source that occurred in random viable networks but not in *E. coli*. The inset shows all 84 secondary carbon sources that are produced by at least one random network viable on glucose.

secondary carbon sources (mean 11±5) that can sustain a two-strain community (Figure 2.3B). Taking all sampled metabolic networks we find 84 secondary carbon sources that are produced by at least one random viable network (26 more than produced by *E. coli*). Most secondary carbon sources are produced by more than one random viable network. Figure 2.3C shows all secondary carbon sources that are produced by any random network, ranked by the fraction of the 500 random viable networks that produce them. Acetate and glycerol, the secondary carbon sources found experimentally when growing *E. coli* on glucose are among the top-ranked carbon sources, with respective ranks of 19 and 13.

When exposed to not just glucose but to each of its primary carbon sources in turn, a random viable metabolic network can produce on average 14±6 secondary carbon sources (ranging from 0 to 26). The number of primary-secondary carbon source pairs varied greatly between networks, ranging from 0 to 1065 (mean 404±236). In total we observe 15685 different primary-secondary carbon source pairs that could serve as the foundation of a stable community in at least one random viable network.

In sum, because even random viable metabolisms show high niche construction and cross-feeding potential, this potential is likely an intrinsic property of metabolic systems.

2.3.6 Niche construction potential in the pan-metabolic network

Our last analysis complements the previous analysis by constructing a pan-metabolic network that contains all metabolic reactions from a known and curated "universe" of metabolic reactions (Methods). For various reasons, such a network could never be realized in any one organism, but it provides another way to inform us which kind of cross feeding interactions are metabolically possible. (Supplementary 2.13 Fig illustrates how this pan-metabolism analysis relates to our previous analysis of random viable networks.) The pan-metabolic network we analyzed comprises 7222 reactions and 5625 metabolites. As in our analysis of random viable metabolic networks, we only allowed those 330 metabolites to enter and leave a cell that can also enter or leave *E. coli*. This focuses our analysis on novel biosynthetic abilities rather than on novel transport as a reason for the production or consumption of novel carbon sources. It also implies that we may underestimate the numbers of primary and secondary carbon sources, and perhaps dramatically so.

The pan-metabolic network harbors 221 metabolites that can be used as primary carbon sources, 41 more than *E. coli*. The minimum number of secondary carbon sources produced per primary carbon source is 85 (supplementary 2.11C Fig), compared to the 54 in *E. coli* (supplementary 2.11A Fig). In total, the pan-metabolic model can produce 109 secondary carbon sources. The total number of primary-secondary carbon source pairs almost doubles relative to *E. coli* (18959 and 9913 pairs for the pan-metabolic and the *E. coli* network, respectively). See table 2.1.

In sum, the numbers of primary-secondary carbon source pairs that can sustain stable communities

is greatest in the pan-metabolic network. In different organisms, different subsets of such pairs may be suitable for cross-feeding induced niche construction.

2.4 Discussion

When an isogenic bacterial population grows in a homogeneous environment with a single nutrient or carbon source, the organisms in the population initially behave similarly and consume this nutrient. They may also excrete by-product metabolites that accumulate in the environment. If they can express the necessary enzymes, they may switch to consume the by-products once most of the initial nutrient is consumed. In such a population, DNA mutations may arise that alter metabolic properties like enzyme activities or expression permanently. As a result, ancestor and mutant strains may compete for the original nutrient, and one of them may eventually be excluded from the population. Alternatively, a mutant may specialize in the consumption of the ancestor's by-products. Our work focuses on this commensal or mutualistic scenario, which can help ancestor and mutant to coexist stably, and thus permanently increase genetic and metabolic diversity.

We searched exhaustively for by-product or secondary carbon sources that can be excreted when a microbial strain grows on some primary carbon source, and that can themselves sustain microbial life. We performed this search with the metabolism of three non closely related organisms: *E. coli*, *B. subtilis* and *S. cerevisiae*; with 500 randomly sampled metabolic networks that we required to be viable on at least glucose and that contained the same number of biochemical reactions as *E. coli*; and with a pan-metabolic network containing 7222 biochemical reactions known to occur in extant organisms. For each of these metabolisms we identified thousands of possible cross-feeding interactions where one strain produces a carbon source that can sustain the other strain. Through a combination of analytical calculations and simulations of the ecological dynamics of two-strain chemostat communities, we demonstrated the existence of 9919 unique cross-feeding niches in *E. coli* alone that can sustain a stable two-strain community. Each niche corresponds to a unique pair of primary and secondary carbon sources. Our observations suggest an enormous potential for population diversification through niche construction and cross feeding.

Although it may seem puzzling that an organism would dispose of metabolites that could advance its own growth, it is not an unusual phenomenon. The causes are multiple, and include membrane leakage, overflow metabolism, genetic mutations, and cells fermenting carbon sources even in the presence of oxygen (Pfeiffer and Bonhoeffer, 2004; Xu et al., 1999; Pfeiffer et al., 2001). In addition to acetate, for example, *E. coli* frequently releases formate, lactate, succinate and ethanol into the environment as a result of fermentation or membrane leakage (Xu et al., 1999). Various microorganisms, including *Escherichia coli*, *Corynebacterium glutamicum*, *Bacillus licheniformis* and *Saccharomyces cerevisiae*, excrete a broad diversity of more than 30 metabolic intermediates and amino acids (Paczia

et al., 2012). Detecting such secondary carbon sources may promote the experimental discovery of new cross-feeding interactions.

Our work differs in various ways from previous studies on microbial metabolic interactions that include competition, commensalism and mutualism (Chiu et al., 2014; Tzamali et al., 2011; Klitgord and Segrè, 2011; Stolyar et al., 2007; Harcombe et al., 2014a; Zelezniak et al., 2015; Zomorrodi and Maranas, 2012; Levy and Borenstein, 2013; Estrela et al., 2012; Hoek et al., 2016) in general, and cross-feeding in particular (Doebeli, 2002; Pfeiffer and Bonhoeffer, 2004; Gudelj et al., 2016, 2007; Porcher et al., 2001; Keymer et al., 2012; Rocabert et al., 2017; McNally and Borenstein, 2017; Hoek and Merks, 2017; Großkopf et al., 2016; Mori et al., 2016). The most closely related studies (McNally and Borenstein, 2017; Chiu et al., 2014; Tzamali et al., 2011) use a metabolic model of *E. coli* to study the various cross-feeding interactions that can emerge in co-culture after single gene knock-out (Tzamali et al., 2011) or extensive gene loss (McNally and Borenstein, 2017). Our work, in contrast, shows that even without such alterations to its reaction complement *E. coli* can create many niches. In addition, we also analyzed other organisms, as well as random viable metabolisms to demonstrate that this niche construction potential is not just a property of *E. coli* or closely related organisms, but a generic property of complex metabolic systems. Other authors have demonstrated that microbes from different species that are cultured together can show new biosynthetic abilities (Chiu et al., 2014). In contrast, our work shows that new niches and stable communities can emerge from within a population of initially identical individuals. And perhaps most importantly, we have not merely reproduced a single experimentally demonstrated niche construction process, but found that metabolic systems can give rise to myriad new niches through cross-feeding.

Our analysis has several limitations. First, we rely on current knowledge about the metabolism of *E. coli*, *B. subtilis*, *S. cerevisiae* and on reactions in the pan-metabolic network. Future research is likely to discover additional reactions in these networks. They may allow the consumption of additional primary carbon sources, or the synthesis of additional secondary carbon sources. In either case, such additional reactions can only increase, not decrease, the niche construction potential of metabolism.

Second, whereas different organisms can import or excrete a different spectrum of molecules, our analysis of random viable networks and the pan-metabolic network allowed only those metabolites to enter or leave a cell that can also enter or leave *E. coli*. Even so, we found thousands of potential niches. Had we opened cellular transport to further molecules, the number of niches would have increased as well and perhaps dramatically so.

Third, we varied only carbon sources. Similar analyses could be conducted for sources of other chemical elements, such as nitrogen or sulfur. Again, the potential for niche construction could only increase in this case, because different sources of a chemical element can facilitate to the production of novel secondary metabolites. In sum, these limitations, when overcome, would strengthen our

conclusion.

Fourth, random viable metabolic networks and the pan-metabolic network may contain thermodynamically infeasible ATP producing cycles (Beard et al., 2002; Price et al., 2002a; Fritzemeier et al., 2017) that can alter biomass growth. For this reason, it would not have been sensible to simulate cross-feeding dynamics for these metabolic networks. However, our analytical calculations show that the conditions for coexistence hold generally and independently of any one metabolism.

Our observations raise the question why the only known cross-feeding polymorphisms that have been detected in *E. coli* chemostats involve acetate and glycerol as secondary carbon sources. One candidate reason is that many other such polymorphisms exist but have not been detected, because currently no systematic screen for cross-feeding interactions exists. Cross-feeding polymorphisms are usually manifest in different colony morphologies on agar plates, and substantial biochemical and genetic work is needed to prove that such polymorphisms result from cross feeding (Helling et al., 1987; Rozen and Lenski, 2000a; Rosenzweig et al., 1994). A second candidate reason is that in many such polymorphisms, one of the strains may constitute a small fraction of community biomass, which would make its detection even harder. For instance, we showed (Figure 2.2D) that half of the secondary carbon sources that *E. coli* can produce in a glucose environment cause a high metabolic cost to the producer strain or little biomass gain to the consumer strains, which leads to an even lower biomass of the consumer strain than for glucose-acetate cross-feeding. Third, perhaps not all cross-feeding polymorphisms we predict can be biologically realized. For example, on some primary carbon sources multiple regulatory mutations may be needed before a strain produces or consumes some of the secondary carbon sources we predict. Even though such combinations of mutations may arise in large populations of bacteria, the respective secondary carbon sources will be less frequently produced than carbon sources for which single mutations suffice. Characterizing the regulatory mutations needed to bring forth specific secondary carbon sources is a complex undertaking that we will focus on in future work.

Our work focuses on bacterial populations, but similar phenomena may occur elsewhere. For example, they may help explain a hallmark of cancer, the metabolic heterogeneity within tumors (DeBerardinis and Chandel, 2016). Many tumors occupy low oxygen-environments, because they grow faster than blood vessels can form. As a result, they synthesize fermentation products like fumarate or succinate (Jain et al., 2002). In addition, even when oxygen is available, tumor cells exhibit the Warburg effect (Warburg, 1956), the fermentation of glucose to lactate. It is possible that these phenomena may help create new nutritional niches that may be colonized by tumor cells.

Like most biodiversity, bacterial diversity may have arisen through repeated adaptive radiations, in which a single lineage rapidly diversifies to occupy multiple ecological niches (Rainey and Travisano, 1998; Maharjan et al., 2006; MacLean et al., 2005; Travisano and Rainey, 2000). Usually, species

created during adaptive radiations are thought to occupy pre-existing niches, but the rapid emergence of extensive cross-feeding in homogeneous environments (Helling et al., 1987; Rozen and Lenski, 2000a; Good et al., 2017) raises the possibility that many niches are constructed during a radiation. That is, when a bacterial population excretes one or more energy-rich by-product metabolites, it creates niches that can be occupied by mutant strains that are well-adapted to these niches. By excreting their own specific metabolites, these strains can then become stepping stones towards further diversification. In this process, the new metabolic niches into which a population radiates are constructed by the population itself. Because any one bacterial strain can excrete a broad spectrum of metabolites, and because our work identified thousands of niches that could sustain stable communities, the potential for such diversification should not be underestimated. We hope that our observations will motivate experimental work that identifies the extent to which this potential is realized.

2.5 Methods

2.5.1 Flux balance analysis (FBA)

Flux balance analysis (FBA) is a computational method to predict metabolic fluxes – the rate at which chemical reactions convert substrates into products – of all reactions in a genome-scale metabolic network (Orth et al., 2010). FBA requires information about the stoichiometry of chemical reactions in a metabolic network. It makes two central assumptions. The first is that cells are in a metabolic steady-state. The second is that cells effectively optimize some metabolic property such as biomass production (growth). Additional constraints can be incorporated into the optimization problem that FBA solves, in order to account for the thermodynamic and enzymatic properties of a network's biochemical reaction. The optimization problem that FBA solves can be formalized as a linear programming problem (Orth et al., 2010; Varma and Palsson, 1994b) in the following way:

$$\begin{aligned} &Maximize\ v_{growth} \\ &s.t.\ Sv = 0 \\ &l_i \leq v_i \leq u_i \end{aligned} \tag{2.1}$$

Here, S is the stoichiometric matrix, a matrix of size $m \times r$ that mathematically describes the stoichiometry of the network's metabolic reactions. The integer m denotes the number of metabolites and r denotes the number of biochemical reactions in the network. These reactions include all known metabolic reactions that take place in an organism, which are called internal reactions. They also include reactions that represent the exchange (import or export) of metabolites with the external environment. Furthermore, they include a biomass growth reaction, which is a "virtual" reaction that reflects in which proportion biomass precursors are incorporated into the biomass of the modeled organ-

ism (Orth et al., 2011, 2010; Varma and Palsson, 1994b). Each entry S_{ij} of the stoichiometric matrix contains the stoichiometric coefficient with which metabolite i participates in reaction j. The vector v is a vector (of size r) that harbors the metabolic flux through each reaction in the network. v_{growth} specifies the flux through the biomass growth reaction. Fluxes through biochemical reactions are restricted by lower and upper bounds that constrain the flux through each reaction in the network. These bounds are given by the variables l and u, respectively, which are vectors of size r. We performed FBA optimization with the GNU Linear Programming Kit (GLPK; http://www.gnu.org/software/glpk).

2.5.2 The dynamics of producer strain P and consumer strain C in a chemostat

Strains P and C will grow at rates μ_P and μ_C, a process that will increase their respective biomasses X_P and X_C. In a chemostat, fresh medium is continuously added and culture is continuously removed at a dilution rate D. Such dilution leads to a decrease in biomass inside the chemostat. Overall, the change in biomass for P and C can be expressed by the following system of ordinary differential equations :

$$\begin{aligned} \frac{dX_P}{dt} &= (\mu_P - D)X_P \\ \frac{dX_C}{dt} &= (\mu_C - D)X_C \end{aligned} \tag{2.2}$$

The concentration of any one metabolite (M) in a chemostat also varies. If a metabolite is present in the fresh medium at concentration M^0, the metabolite's concentration will increase at a rate DM^0 as a result of fresh medium continually being added to the chemostat. In addition, the metabolite's concentration will also increase if it is produced by strain P (with flux $J^{out}_{M,P}$) or C (with flux $J^{out}_{M,C}$). The total rate at which M is produced will then equal $J^{out}_{M,P}X_P\Delta t + J^{out}_{M,C}X_C\Delta t$. Conversely, the concentration of M will decrease due to removal of old medium at a rate DM, and possibly also due to consumption by P (with flux $J^{in}_{M,P}$) and C (with flux $J^{in}_{M,C}$), at a total rate $J^{in}_{M,P}X_P\Delta t + J^{in}_{M,C}X_C\Delta t$. We denote the net flux of metabolite M as $J_M = J^{out}_M - J^{in}_M$, i.e., which results in a positive net flux J_M if the metabolite is produced and negative otherwise. Overall, the change in concentration for each metabolite present in the chemostat is then described by the differential equation

$$\frac{dM}{dt} = D(M^0 - M) + J_{M,P}X_P + J_{M,C}X_C \tag{2.3}$$

We performed FBA to compute instantaneous growth rates (μ_P and μ_C) in h^{-1}, as well as consumption and excretion fluxes of each metabolite by strains P and C ($J_{M,P}$ and $J_{M,C}$, in $mmol\, gDW^{-1}\, h^{-1}$). We used the values thus computed in dynamic FBA (Varma and Palsson, 1994b) to determine the changing amounts of biomass (expressed in gDW/l) of our microbial strains, as well as the abundance of all metabolites (in mM), nutrients, and waste products in our simulated chemostat.

2.5.3 Simulating chemostat dynamics with dynamic FBA (dFBA)

Dynamic FBA (dFBA) (Varma and Palsson, 1994b) is an FBA-based method to describe the temporal growth dynamics of microbes and how this dynamics affects the microbes' chemical environment. It has been used, for example, to describe chemical growth and by-product secretion of *E. coli* in batch and fed-batch cultures (Varma and Palsson, 1994b), to study the dynamics of a two-species microbial ecosystem in batch culture (Chiu et al., 2014) and to simulate the growth and metabolic dynamics of microbes in time and space (Harcombe et al., 2013).

Briefly, dFBA starts from some initial time point and performs Flux Balance Analysis (FBA) iteratively at each time point during a given time interval to compute the maximally possible growth rate for each strain in the chemostat environment. As microbial strains grow, they consume nutrients and excrete waste products (including, possibly, secondary carbon sources) and thus change their growth environments. Dynamic FBA takes these changes into account by computing the chemical composition of the environment at each time point. In doing so, dynamic FBA predicts how the biomass of bacterial strains and the chemical composition of the environment can change over time.

We used dFBA to predict the temporal behavior of a microbial population composed of acetate producer strain P and consumer strains C in a chemostat. We next describe in detail how we used dFBA to simulate population growth in a glucose-limited minimal medium. We note that our procedure can be applied to any other carbon source by substituting glucose with the desired carbon source.

Our simulations used the following parameters and initial conditions. We chose a dilution rate of $D = 0.2\ h^{-1}$ to mirror conditions from previous experiments that had identified cross-feeding interactions (Helling et al., 1987). We set the glucose concentration in the fresh medium to 1 mM, close to the 0.7 mM used in (Helling et al., 1987). We assumed that ammonium, calcium, chloride, cobalt, copper, iron, magnesium, manganese, molybdate, nickel, oxygen, phosphate, potassium, protons, sodium, sulphate and zinc are present in non-limiting amounts. The initial concentrations of all metabolites in the chemostat are identical to those of the fresh medium. Unless otherwise stated, we initialized the chemostat with 0.01 gDW/l of strain P and 0.001 gDW/l of strain C which corresponds to an overall cell density of approximately 10^7 $cells/ml$ (Milo et al., 2009; Sezonov et al., 2007; Ren et al., 2013). We chose these initial biomass values arbitrarily, except that their unequal values are well-suited to ask whether the consumer strain C, when introduced in small amounts into a culture of strain P, can invade the culture. However, we also show that changing initial biomass values, the ratio of the biomass values, and the time of introduction of C into the chemostat have no effect on the biomass of P and C once steady-state is reached (supplementary 2.4 Fig).

After initializing our simulations, using these parameters, we discretized time into short intervals of 0.1 h and performed dynamic FBA (Varma and Palsson, 1994b) by iterating the following three steps (described in more detail below): calculation of maximum nutrient uptake rates, FBA, and calculation

of environmental composition.

1. *Calculation of maximum uptake rates.* The uptake of a nutrient by an organism is limited by two factors: the capacity to transport the nutrient across the cell wall (transport limitation) and the availability of the nutrient in the environment (nutrient availability limitation). To determine the nutrient transport limit (for a nutrient at concentration M), we assumed Michaelis-Menten kinetics ($V_{max}\ M/(k_M + M)$) with parameter values set to V_{max} =20 $mmol\, gDW^{-1}\, h^{-1}$ and k_M =0.05 mM. These parameters are based on data in the Brenda enzyme database (Scheer et al., 2011; Bar-Even et al., 2011) and have been used in related analyses (Chiu et al., 2014).

At the beginning of a simulation, the nutrient concentration is high and the biomasses of P and C are low. Therefore, nutrient consumption is initially limited by transport. As biomass grows the nutrient begins to be scarce and nutrient availability rather than transport become limiting for nutrient consumption. In other words, the transport limit shapes the transient biomass dynamics but the availability limit determines the steady-state biomass. This also means that V_{max} and k_M can vary over a wide range without affecting the steady state. In supplementary 2.4 Fig we demonstrate the chemostat dynamics for various values of V_{max} and k_M to exemplify how these parameters modify the transient biomass dynamics, but not the steady state.

To determine the nutrient availability limit, we divided the nutrient concentration M by the nutrient consuming biomass. This biomass depends on how much nutrients the strains consumed in the immediately past (according to equations 2.4 and 2.5 in supplementary 2.6.2 Text). We describe and justify our procedure to calculate the nutrient availability limit, which differs from that of some other authors, in depth in the supplementary material (supplementary 2.6.1 Text).

Once we had determined a strain's transport limit for a nutrient and the nutrient's availability limit, we set the uptake rate of the nutrient to the minimum of both, which ensures that organisms do not consume more of a nutrient than is physiologically feasible and available to them.

2. *FBA.* Once we had calculated maximum uptake rates of nutrients as just described, we performed FBA for each strain independently. The calculation yielded growth rate values (μ_P and μ_C) for both strains, as well as consumption or excretion rates of each metabolite M for both strains ($J_{M,P}$ and $J_{M,C}$).

3. *Calculation of environmental composition.* With the results of FBA in hand, we used Euler's method (John C. Butcher, 2003) to determine the environmental change caused by nutrient consumption, waste production, and biomass growth. We did so in accordance to equations 2.2 and 2.3, using the conditions from the beginning of this section and a time increment of 0.1 h.

We repeated these three steps until at most 1000 h (104 time steps) had elapsed or until the chemostat had reached steady state. We assumed that steady state had been reached if the standard deviation of growth rates determined over 50 consecutive time steps was smaller than 10^{-5} for both

strains. We carried out these simulations using MATLAB (Mathworks Inc.).

2.5.4 Search for primary and secondary carbon sources

We searched for all metabolites that could serve as carbon sources in the following way. To identify primary carbon sources, we first considered all metabolites in the *E. coli* model iJO1366 (Orth et al., 2011) a candidate primary carbon source, if it contained at least one carbon atom and if *E. coli* had an exchange reaction for this carbon source. Second, we used FBA to determine *E. coli*'s maximal biomass production when each of these primary carbon sources was available as the sole carbon source. (We assumed that ammonium, calcium, chloride, cobalt, copper, iron, magnesium, manganese, molybdate, nickel, oxygen, phosphate, potassium, protons, sodium, sulphate and zinc can be consumed without constraints). Third, if any one carbon source was able to sustain non-zero biomass production, we considered it an actual primary carbon source. Here and below, we viewed only biomass production fluxes above 10^{-5} $mmol\,gDW^{-1}\,h^{-1}$ as being different from zero.

Our approach identified 180 primary carbon sources (Figure 2.2A). On about half of these carbon sources, growth of *E. coli* has been demonstrated experimentally (AbuOun et al., 2009; Feist et al., 2007). No experimental data is available for multiple other carbon sources. Metabolic reconstruction errors may account for the discrepancies between computational predictions and experimental observations for some other carbon sources, but at least for the well-studied *E. coli*, they may be a minor cause compared to regulatory constraints that are not incorporated by most genome-scale models analyzed with FBA (Feist et al., 2007). Such regulatory constraints, where enzymes are encoded by a genome but are not expressed when needed, can be easily broken. That is, even on the short time scales of laboratory evolution, microbial populations can adapt to grow on a novel carbon source in accordance with FBA predictions (Ibarra et al., 2002). Because regulatory evolution can occur during the long-term cultivation of *E. coli* that we model, we assume that regulatory constraints can be by-passed, and thus use all 180 primary carbon sources on which FBA predicts growth in our analyses.

We considered a metabolite a secondary carbon source if (i) it can serve as a primary carbon source and (ii) if it can be produced as a metabolic by-product when another metabolite serves as a primary carbon source. The first condition ensures that the metabolite can sustain growth of a strain consuming it, and the second condition ensures that the metabolite can be produced. Note that all primary carbon sources are potential secondary carbon sources, but only some of them may be produced as metabolic by-products in a given environment. Most importantly, whether a carbon source is produced depends on the available primary carbon source. To identify actual secondary carbon sources and distinguish them from potential ones, we iterated through all pairs of primary carbon sources and potential secondary carbon sources, and performed FBA. More specifically, we used the primary carbon source as the sole carbon source (uptake rate: 10 $mmol\,gDW^{-1}\,h^{-1}$), maximized

the production of the potential secondary carbon source, and constrained biomass production in FBA to be greater than zero. If the potential secondary carbon source could be produced at a rate greater than zero under this constraint, we considered the carbon source an actual secondary carbon source.

We used the same method described in the previous paragraphs to search for primary and secondary carbon sources in the genome scale metabolic networks of *B. subtilis* (model iYO844 (You-Kwan Oh et al., 2007)) and *S. cerevisiae* (model iMM904, (Mo et al., 2009)), modifying only the chemical environment. Specifically, for *B. subtilis* we used an environment composed of ammonium, calcium, carbon dioxide, iron, magnesium, oxygen, phosphate, potassium, protons, sodium and sulphate. We constrained the ammonium, phosphate and sulphate uptake rates of *B. subtilis* to a maximum of 5 $mmol\, gDW^{-1}\, h^{-1}$. For *S. cerevisiae* , we used a medium consisting of ammonium, iron, oxygen, phosphate, potassium, protons, sodium and sulphate, and constrained the oxygen uptake rate to a maximum of 2 mmol gDW-1h-1. We obtained all three metabolic models (iJO1366, iYO844 and iMM904) from the BiGG Database (Schellenberger et al., 2010).

2.5.5 Pan-metabolic network

The pan-metabolic network is a network containing all metabolic reactions with well-defined stoichiometry that are known to take place in some organism. For our analysis, we extended a previously used pan-metabolic network comprising 5484 metabolites and 6892 reactions (Barve and Wagner, 2013) by adding the 141 metabolites and 330 reactions from *E. coli* iJO1366 that were not already present in this network. This amended pan-metabolic network includes 5625 metabolites and 7222 reactions.

We found that 3070 reactions (43%) in the pan-metabolic network are unconditionally blocked (Burgard et al., 2004). That is, they cannot carry non-zero flux without violating FBA's steady state assumption when all metabolites to which *E. coli* is permeable can freely enter and leave the cell. We note that if more metabolites where allowed to enter and leave the pan-metabolic network the number of blocked reactions would decrease. For reference, in the well-curated iJO1366 model 227 reactions (9%) are unconditionally blocked. In terms of absolute numbers, the pan-metabolic network contains 1569 more reactions that can carry nonzero flux than the iJO1366 model.

Genome scale metabolic network reconstructions often contain spurious energy producing cycles that violate the second low of thermodynamic (Beard et al., 2002; Price et al., 2002a; Fritzemeier et al., 2017). Unless removed from the network, these cycles can spuriously increase biomass production. (For example, removal of these cycles causes an approximately 25% reduction of biomass production in 92% of the networks analyzed in (Fritzemeier et al., 2017)). The pan-metabolic network that we use contains reactions that create spurious ATP producing cycles, allowing ATP to be produced even in the absence of nutrients. However, because 67 metabolites in addition to ATP must be produced for biomass growth, biomass cannot be produced in the absence of nutrients. We emphasize that we

evaluated the pan-metabolic network's viability on specific carbon sources and its ability to produce secondary carbon sources without any quantitative evaluation of fluxes, for which spurious cycles might be a problem.

2.5.6 Random viable metabolic networks

We wanted to study the niche construction capacity not only of *E. coli*, but of multiple networks of similar complexity that do not share *E. coli*'s or any other organism's evolutionary history. Any one metabolic network can be thought of as a subset of reactions drawn from the set of all metabolic reactions feasible in a living organism, i.e., the pan-metabolic network. In the enormous space of all possible metabolic networks only a tiny fraction is viable on any one carbon source, i.e., they can produce biomass when this carbon source is the sole carbon source. We focused on such viable networks, and to sample them from the space of such networks, we used a technique based on Markov Chain Monte Carlo (MCMC) sampling (Samal et al., 2010; Rodrigues and Wagner, 2009), which samples networks during long random walks in the space of all metabolic networks of a given size. The statistical theory behind MCMC sampling (Brooks, 1998) shows that its random walks are ergodic, i.e., roughly speaking, they are equally likely to visit all metabolic networks in a connected region of the space of such networks. In previous work, we have shown that in the space of all possible networks, networks viable on a specific carbon source form indeed a subset connected by single reaction changes (Barve et al., 2014). One requirement of the method is that random walks have a sufficiently long burn-in period to ensure that any "memory" of the starting network of such a random walk has decayed. In previous work (Samal et al., 2010), we determined that a burn-in period of 5000 reaction changes is sufficient for this purpose. When this requirement is met, the method essentially ensures that the sampled networks contain a random complement of reactions, with no similarity to the starting network in excess of that required for viability on a specific carbon source.

The method starts from an initial network, which we chose as a network that is viable on glucose as a sole source of carbon and that has the same number of reactions (2583) as *E. coli*. (The computational cost of MCMC sampling prevented us from exploring other primary carbon sources.)To create this initial network, we first performed Flux Balance Analysis on the pan-metabolic network, with glucose as the only source of carbon. Of all reactions in the pan-metabolic network, 1263 reactions showed non-zero flux and were included in the initial network, which ensured viability on glucose. We chose the remaining (1320) reactions needed to arrive at an equal number of reactions as *E. coli* at random from the pan-metabolic network. From this initial network the MCMC method creates a long sequence of modified networks. In our implementation of the method, a new network is created by a reaction swap, in which a reaction from the existing network is randomly chosen for deletion, while a randomly chosen reaction (that is not yet present in the network) from the pan-metabolic network is added to

the network. If the network remains viable after this reaction swap, the swap is accepted, and the network is modified with a further reaction swap. In contrast, if the reaction swap disrupts viability on glucose, the swap is rejected and a new swap is tried. Modifying metabolic networks through reaction swaps ensures that the number of reactions in the network remains constant (and equal to the number of reactions in the initial network). As the number of reaction swaps increases, the number of reactions that the altered networks share with the initial network becomes smaller and smaller, until the complement of reactions has become effectively randomized after 5000 successful swaps (Burgard et al., 2004). We stored such a randomized network (which is still viable on glucose) for further analysis after 5000 successful swaps.

We performed 500 independent such random walks, thus creating 500 metabolic networks all viable on glucose and containing as many reactions as the network of *E. coli* does. Each of them is the end point of a sequence of 5000 successful reaction swaps. This procedure is very time consuming, and to accelerate it, we first determined the reactions that are essential for growth on glucose in the pan-metabolic network, and did not subject these (169) reactions to deletions.

We note that we did not alter the exchange reactions of the starting network, which ensures that in the randomized networks the same metabolites can be exchanged with the environment as in *E. coli*. We also note that random networks contain a large number of reactions (1197±36) that cannot carry non-zero flux in any of the environments we consider, i.e., they are unconditionally blocked. Even though all networks analyzed in this work contain unconditionally blocked reactions, the numbers observed for the random viable networks are especially large.

Because we created random networks by sampling reactions from the pan-metabolic network, these networks may also contain spurious cycles (Beard et al., 2002; Price et al., 2002a; Fritzemeier et al., 2017). For this reason we analyzed them similarly to the pan-metabolic network, evaluating only viability on specific carbon sources and the ability to produce secondary carbon sources, without any quantitative evaluation of fluxes which are most affected by spurious cycles.

We implemented our sampling procedure in python, using the cobrapy package (Ebrahim et al., 2013) to perform flux balance analysis to check for viability on glucose of the networks created after each reaction swap.

2.6 Supplementary material

2.6.1 Calculating the nutrient availability limit

In this section, we describe how to determine the nutrient availability limit for each environmental nutrient, and for each strain in our simulated chemostat. This maximal availability is required to calculate the maximum nutrient uptake rate, which FBA requires to determine the maximal biomass

production rate.

The two strains in our chemostat do not necessarily compete for any one nutrient. For example, a nutrient may be available in sufficiently high concentrations that it does not limit growth, or only one of the strains may be able to use the nutrient for growth. In such cases it is straightforward to determine the nutrient availability limit: It is calculated by dividing the total concentration of the nutrient by the biomass of the strains consuming the nutrient.

In contrast, it is less straightforward to determine a nutrient's availability limit when the strains compete for this nutrient. Some authors (Harcombe et al., 2014a) select randomly to which strain they allocate the whole amount of nutrient existing in the environment at a given time. The remaining nutrients can be consumed by the other strain. This strategy is reasonable but it may alter the ecological dynamics of the strains depending on which strain is selected for nutrient allocation at each time step of the simulation. Other authors (Chiu et al., 2014) divide the nutrient's concentration by the total biomass (adding the biomass of all the strains) and allocate equal amounts of nutrient to be consumed by each strain. Although this approach is also reasonable, it eliminates frequency dependency effects that can arise when a strain with small biomass consumes a nutrient that is not consumed by the other strain. Specifically, consider two strains P and C that are present in the chemostat. Strain P is abundant whereas C is rare. Both can in principle consume the nutrient m (they express the necessary enzymes to do so), but when strain P grows its biomass at the maximally possible rate, as predicted by FBA, it does not consume m, perhaps because maximal biomass growth is achieved with other available nutrients. If we calculated the nutrient availability limit for C by simply dividing the concentration of m by the sum of the biomass values of P and C, we would allow little of m to be consumed by C. In consequence, C's growth rate would remain low, and most of m would remain unutilized in the chemostat, because it is not consumed by the much more abundant strain P.

To overcome undesirable artifacts like this, we pursued the following, third strategy, which is more appropriate for our purpose. At each time step t_i, both strains are allowed to consume a given nutrient, but we make the amount allocated to each strain dependent on their consumption at the previous time step t_{i-1}. By doing so, we encourage continuity with respect to the strains' nutrient consumption. With two strains (P and C), there are four possibilities regarding nutrient consumption at time t_{i-1}: 1) both strains consumed the nutrient, 2) only strain P did, 3) only strain C did, 4) none of the strains did. In case 1) both strains consumed the nutrient at time t_{i-1}, so we simply calculate the nutrient availability limit by dividing the nutrient's concentration by the total biomass. In case 2), where only strain P consumed the nutrient, we calculate the nutrient availability limit for P by dividing the metabolite's concentration by the biomass of P. At time step t_i, since the environment has changed with respect to time t_{i-1} it may have become beneficial for C to consume the nutrient of interest,

i.e., doing so may help increase the growth rate of C. To allow for this possibility, we calculate the nutrient availability limit for C by dividing the nutrient's concentration by the total biomass of both strains. For case 3) the nutrient availability limitcalculates analogously to case 2, but with strain identities reversed. In the last scenario 4), where none of the strains had consumed the nutrient at the previous time step, we permit both strains to consume the available nutrient. In this case, the nutrient availability limit for each strain is obtained by dividing the nutrient concentration by the strain's biomass. In theory, under scenario 2, 3 or 4, either strain may consume all nutrient at time t_i, and total nutrient consumption may thus exceed the amount of available nutrient. In practice, this will not occur if the time steps are small enough, and for our time step size ($0.1\ h^{-1}$) it never occurred. Note that in scenario 2, 3 or 4, the availability limit of the nutrient may differ between the strains.

Overall, this strategy provides continuity in the nutrient consumption pattern of the two strains from one time step to the next, while still permitting change in this pattern. We can formalize the calculation of the nutrient availability limit as follows. For strain P at time t_i the availability limit of a nutrient (metabolite) with concentration M calculates as

$$M/((X_P + f_{M,C}(t_{i-1})X_C)\Delta t) \tag{2.4}$$

Here, $f_{M,C}(t_{i-1})$ can take values of one and zero, corresponding to the situation where strain C consumed or did not consume the nutrient at time t_{i-1}, respectively. The variables X_P and X_C denote the biomass values of strains P and C, respectively. With analogous notation, the availability limit of this nutrient for strain C at time t_i calculates as

$$M/((X_C + f_{M,P}(t_{i-1})X_P)\Delta t) \tag{2.5}$$

2.6.2 Biomass yield, maximal production and cost

We define the biomass yield α_m of a nutrient or metabolite as the growth rate a metabolism can achieve per consumed unit of flux of the metabolite. To compute this yield in E. coli for any one metabolite, we allowed an uptake rate (J_m^{in}) of 10 $mmol\, gDW^{-1}\, h^{-1}$ of the metabolite, and maximized biomass production with Flux Balance Analysis. We then reported the yield as the maximal growth rate obtained per unit flux of metabolite consumed.

A second quantity relevant to our analysis is the maximal production rate of a metabolite p_m^{max}. We define this rate as the maximal rate of a metabolite's production that allows the organism producing the metabolite to grow at a rate that is identical to the chemostat's dilution rate D. In other words, it is the highest metabolite production rate at which the organism will not get flushed out of the chemostat over time. For the calculation of this rate, we performed FBA to maximize the production of the desired metabolite, while constraining the growth rate to the value of the dilution rate D.

A third relevant quantity is the metabolic cost of producing a metabolite. To calculate this cost, we computed the maximal biomass growth rate with FBA twice, first without any additional constraints, and then by requiring that the focal metabolite is produced at a rate of 1 $mmol\, gDW^{-1}\, h^{-1}$. We quantify the cost of producing the metabolite as the reduction in growth rate when the metabolite is produced at this rate.

We note that the cost and maximal production rate of a metabolite are closely related and show an inverse relationship. The greater the cost, the smaller the maximum rate that a metabolite can be produced at a given biomass growth rate (supplementary 2.5 Fig).

2.6.3 The limits of coexistence when strains compete for the primary carbon source

In the main text we explored a simple scenario in which the strain C consuming a secondary carbon source cannot metabolize the primary carbon source ($c_{glc,C} = 0$). This scenario leads to coexistence of strains P and C. In this section we explore the consequences of a more realistic but complex scenario that emerges when strain C can consume not only the secondary carbon source, but also the primary carbon source (Figure 2.1A of main text with $c_{glc,C} > 0$). In this scenario, the strains compete for the primary carbon source. To identify the conditions for their coexistence, we varied the producer's acetate synthesis rate between 0 and 100% of the maximally possible value $p^{max}_{ac,P}$ beyond which the strain would eventually be flushed out of the chemostat. To model the consumer strain's dynamics, we first used FBA to determine the minimal glucose consumption rate $c^{min}_{glc,C}$ C would need to persist in the chemostat by itself (in the absence of the other strain), if glucose were the only carbon source. This minimal glucose consumption rate is 2.04 $mmol\, gDW^{-1}\, h^{-1}$ for our dilution rate of D=0.02 h^{-1}. We then varied C's glucose consumption rate $c_{glc,C}$ between zero and a value larger than this minimum, at which C can persist in the chemostat by consuming only glucose.

Supplementary 2.5 Fig shows the steady-state composition of the chemostat for different values of glucose consumption by C (horizontal axis), expressed as a percentage of the minimal glucose consumption rate $c^{min}_{glc,C}$ under which strain C can persist on glucose alone, and acetate production by P (vertical axis), expressed as a percentage of the maximum acetate production rate $p^{max}_{ac,P}$ under which strain P persists. Depending on these rates, C alone may persist (dark grey), P alone may persist (light grey), neither strain may persist (black), or both strains may persist (coexistence, colored).

When C's glucose consumption exceeds $c^{min}_{glc,C}$, such that it could persist on glucose alone, then only two persistence outcomes are possible. First, only C may persist (supplementary 2.5 Fig, dark grey). This occurs if P produces acetate, which inflicts a metabolic cost on P. This cost reduces P's growth rate, and thus causes a growth disadvantage relative to C, which leads to P's eventual extinction. Second, the two strains coexist but are metabolically indistinguishable (supplementary 2.5

Fig bright yellow). This occurs when P produces no acetate, such that C uses the same resources as P for growth (supplementary 2.6 Fig). Since the two strains harbor identical metabolic networks, they metabolize glucose in the same way, respiring it completely to carbon dioxide, and are thus metabolically indistinguishable.

When C's glucose consumption is insufficient for its persistence ($c_{glc,C} < c_(glc, C)^min$) three outcomes are possible. First, if P produces no acetate, P persists but C goes extinct (supplementary 2.5 Fig, light grey), because C consumes insufficient glucose for its persistence and because no acetate is available to supplement its glucose consumption. Second, if P produces acetate at the maximal rate $p^{max}_{ac,P}$, P and C go extinct. P goes extinct because it has to produce more acetate than it is capable of while persisting. After P's extinction, the acetate it produced will eventually disappear from the chemostat, which leads to the extinction of C, because C cannot persist on glucose alone. For intermediate acetate production ($0 < p_{ac,P} < p^{max}_{ac,P}$) the two strains coexist stably and cross-feed.

In the experiments that motivated this work (Rosenzweig et al., 1994) the steady state biomass of the producer strain was reported to be approximately nine times that of the consumer strain. We predict this ratio when P produces little acetate ($\sim 1\%$ of the maximal acetate production rate $p^{maz}_{ac,P}$) and C consumes glucose at a rate of up to $\sim$70% of the minimum rate $c^{min}_{glc,C}$ needed for growth on glucose alone. To our knowledge, the maximal acetate production $p^{max}_{ac,P}$ and minimal glucose consumption rates have not been measured experimentally. However, our prediction is reasonable given the short divergence times and high metabolic similarity between strains, because it requires only that the producer strain produces very little acetate and the consumer strain consumes high amounts of glucose.

2.6.4 Analytical analysis of the limits for coexistence

In supplementary 2.5 and 2.6 Figs we observed that for some combination of acetate production by P ($p_{ac,P}$) and glucose consumption by C ($c_{glc,C}$), two metabolically distinguishable strains cannot coexist. In this section, we analyze the conditions for coexistence analytically to demonstrate the generality of our observations. In the following, we use notation specific to glucose and acetate, but we emphasize that our conclusions are independent of the primary and secondary carbon source considered.

At a metabolic steady state (ss), if P and C strains are to coexist, the growth rates of the producer strain P and the consumer strain C must be equal to the dilution rate of the chemostat. In mathematical terms (see also equation 2.2 in the main text):

$$\mu^{ss}_P = \mu^{ss}_C = D \tag{2.6}$$

The producer strain P consumes glucose at some rate $J^{in}_{glc,P}$. Part of this glucose flux is used for acetate production and the rest is used for growth. The amount used for acetate production will

depend on $p_{ac,P}$, i.e., the acetate flux to be produced. It will also depend on the 'cost' of acetate production, β^{ac}_{glc}, which we define as the growth reduction observed per unit of acetate flux produced, when a strain grows on glucose. The growth rate for the producer strain can then be expressed as

$$\mu_P = \alpha_{glc}(J^{in}_{glc,P} - \beta^{ac}_{glc} p_{ac,P}) \tag{2.7}$$

In other words, strain P consumes glucose at a rate $J^{in}_{glc,P}$, and part of the consumed glucose ($\beta^{ac}_{glc} p_{ac,P}$) is used for acetate production. The remainder ($J^{in}_{glc,P} - \beta^{ac}_{glc} p_{ac,P}$) is used for growth, and the growth rate can be calculated by taking into account the biomass yield of glucose (α_{glc}) (Supplementary 2.6.2 Text).

Similarly, the growth rate of the consumer strain depends on the amount of glucose and acetate consumed and their respective biomass yields (α_{glc} and α_{ac}) in the following way:

$$\mu_C = \alpha_{glc} J^{in}_{glc,C} + \alpha_{ac} J^{in}_{ac,C} \tag{2.8}$$

In metabolic steady state, the dilution flux D must be equal to the biomass production for any strain persisting in the chemostat (equation 2.6), which allows us to derive a simple condition for the persistence of consumer strain C in the presence of strain P. Specifically, by combining equation 2.6 and 2.8, we obtain $D = \alpha_{glc} J^{in,ss}_{glc,C} + \alpha_{ac} J^{in,ss}_{ac,C}$ and therefore $J^{in,ss}_{glc,C} = (D - \alpha_{ac} J^{in,ss}_{ac,C})/\alpha_{glc}$. By definition, $J^{in,ss}_{ac,C} > 0$ must hold for any acetate consumer strain C, and since α_{ac} is always positive, we obtain

$$J^{in,ss}_{glc,C} < D/\alpha_{glc} \tag{2.9}$$

This expression indicates that as long as the glucose consumption rate of strain C is lower than the dilution rate divided by the biomass yield of glucose, the metabolically distinguishable strains P and C can coexist. Higher dilution rates D will permit higher glucose consumptions by C simply because to sustain growth at higher dilution rate, more carbon must be consumed (either as primary or secondary carbon source). The same reasoning explains why consuming a primary carbon source with lower biomass yield than glucose permits higher primary carbon source consumption by C for coexistence. That is, a reduced biomass yield of the primary carbon source leads to a lower growth rate per unit flux, such that a higher rate of consuming C becomes acceptable for coexistence. (We note that this entire analysis assumes that some acetate exists in the medium to be consumed by strain C, i.e., that $0 \leqslant p_{ac,P} \leqslant p^{max}_{ac,P}$.

We have determined the biomass yield of glucose in the *E. coli* metabolism (iJO1366) as $\alpha_{glc} = 0.098$. With a dilution rate of D=0.2 used throughout this paper, equation 2.9 gives a maximum

glucose consumption rate by C of 2.04 $mmol\, gDW^{-1}\, h^{-1}$ at which two metabolically distinguishable producer and consumer strains can coexist. This analytically derived rate explains the change from coexistence to non-coexistence we observed in our simulations (supplementary 2.5 and 2.6 Figs) as the glucose consumption rate of strain C ($c_{glc,C}$) increases from $c_{glc,C} = 2.02$ (99%) to $c_{glc,C} = 2.04$ (100%).

In addition to these considerations, we can also combine equations 2.6, 2.7 and 2.9 to obtain a value for their relative rate of glucose consumption that the strains must fulfill for coexistence. Expressed as a ratio

$$\frac{J^{in,ss}_{glc,C}}{J^{in,ss}_{glc,P}} < 1 - \frac{\beta^{ac}_{glc} p_{ac,P}}{J^{in,ss}_{glc,P}} \tag{2.10}$$

The inequality shows that although P and C can both consume glucose, coexistence requires that their consumption rates are unequal. Specifically, C's glucose consumption rate must be lower than P's glucose consumption rate, and how much lower is given by the cost of acetate production and the amount of acetate $p_{ac,P}$ to be produced (supplementary 2.6 Fig). We emphasize again that the above analyses, in particular equations 2.9 and 2.10, are not specific to glucose and acetate, but apply to any primary and secondary carbon source pair.

2.6.5 Supplementary figures

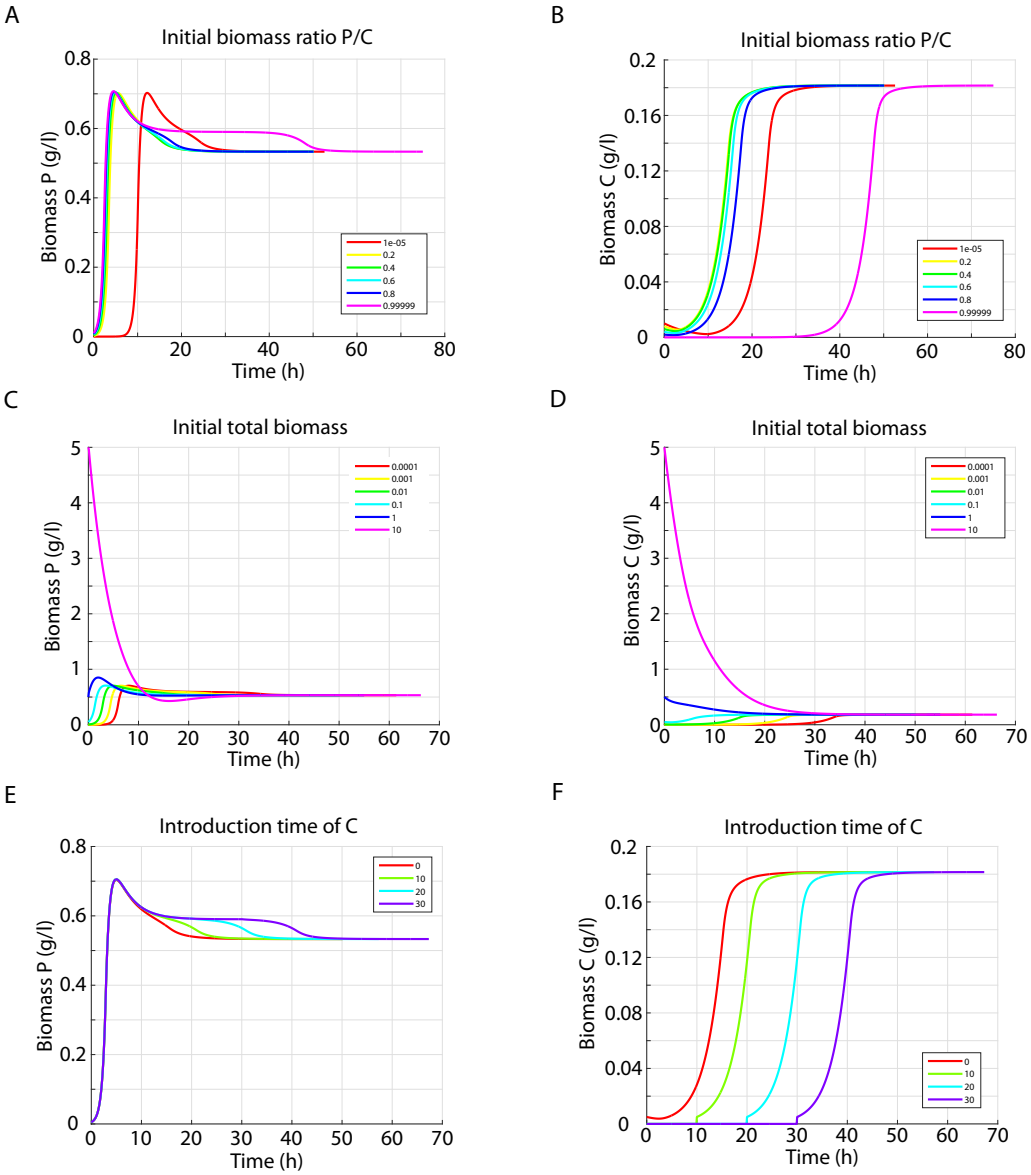


Figure 2.4: Dynamics in the chemostat for various initial conditions. Horizontal axes in all panels indicate time in hours. Panels (A) and (B) show the biomass values of P and C respectively (vertical axes), as a function of time, while changing the initial biomass ratio of P and C (see color legends in both panels). Panels (C) and (D) show the biomass values of P and C respectively, as a function of time when changing the total initial biomass, while maintaining the ratio of P to C biomass at a constant value of 0.5 (see color legends in both panels). Panels (E) and (F) show the biomass values of P and C, respectively, as a function of time for two scenarios. In the first, the chemostat is initiated with equal amounts of P and C biomass at time zero (red line). In the second, the chemostat is initiated just with P, and C is introduced at various times in an amount equal to that of P at time zero (see color legend).The chemostat composition and all other parameters used in these simulations are identical to those described in the main text (Section "Simulating chemostat dynamics with dynamic FBA" in methods).

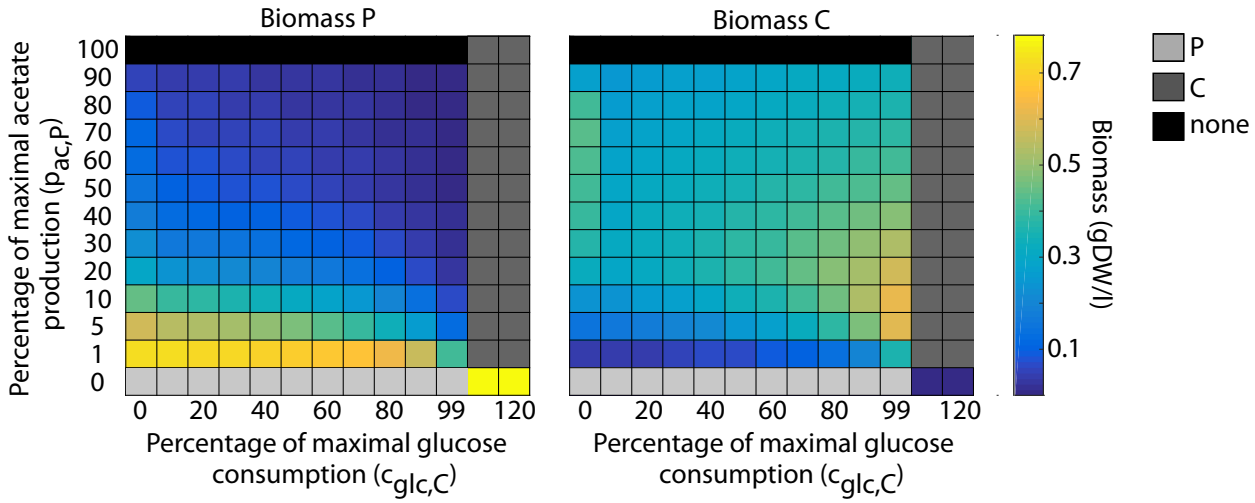


Figure 2.5: Steady state biomass of P (left panel) and of C (right panel) in the chemostat. The amounts of biomass are shown as a function of P's acetate production rate (vertical axis) and of C's glucose consumption rate (horizontal axis), which are expressed as percentages of the maximal acetate production rates and glucose consumption rates that permit coexistence of metabolically distinguishable strains P and C. Note the nonlinear scale used at low values of $p_{ac,P}$ (vertical axis). The amount of biomass is indicated by a color gradient (see color legend) in the region where the two strains can coexist. Note that for parameter combinations where no acetate is produced and where glucose consumption is higher than 99% of $c_{glc,C}$, both strains coexist but are metabolically indistinguishable, because both completely respire glucose to carbon dioxide. Areas (parameter combinations) where coexistence is not possible are shown in light grey (only P persists), dark grey (only C persists) and black (neither strain persists).

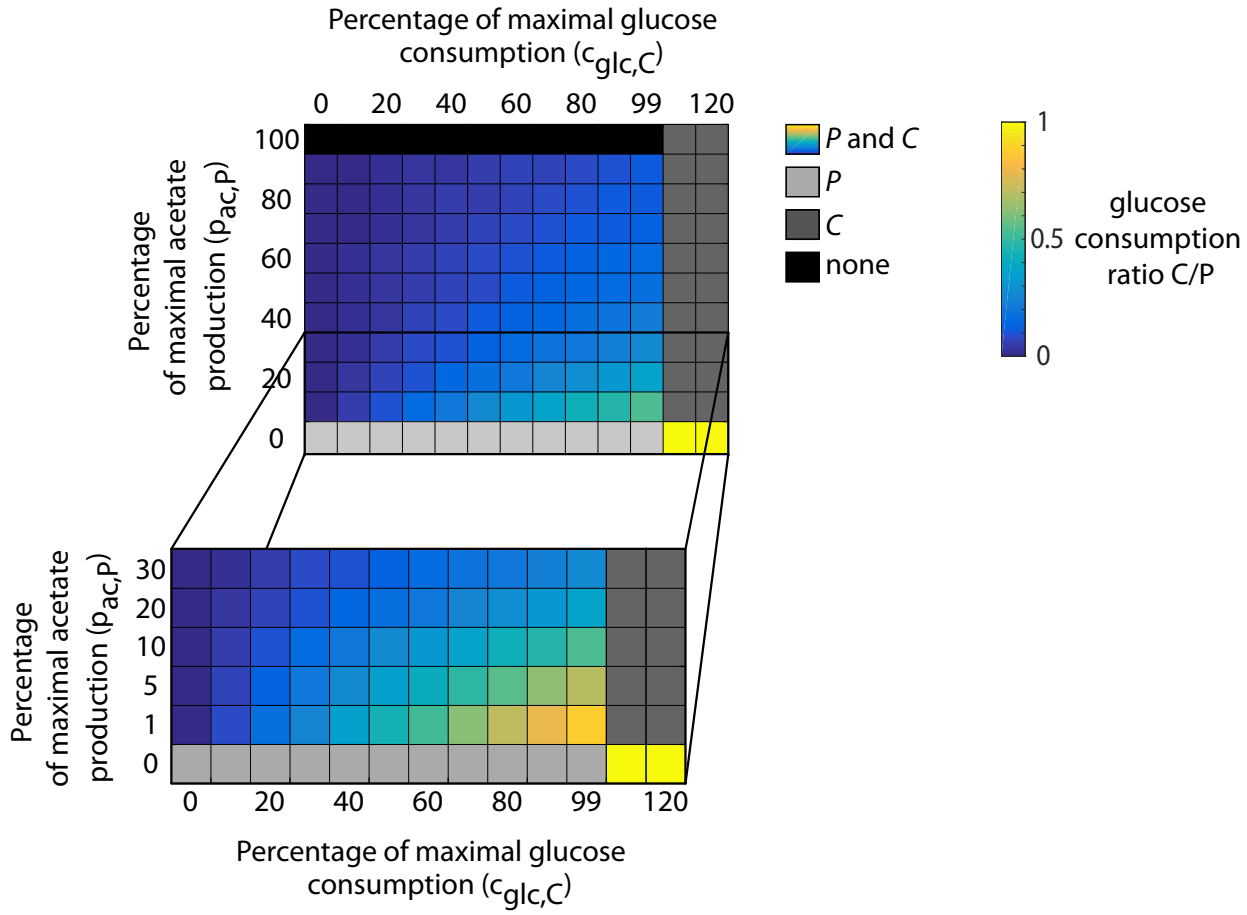


Figure 2.6: Steady state glucose consumption ratio. The figure shows the ratio of glucose consumption rates of strains C and P, as a function of P's acetate production rate (vertical axis) and of C's glucose consumption rate (horizontal axis), expressed as percentages of the maximal acetate production rates $p_{ac,P}$ and glucose consumption rates $c_{glc,C}$ that permit coexistence of metabolically distinguishable strains P and C. As in supplementary 2.5 Fig, depending on the values of $c_{glc,C}$ and $p_{ac,P}$, the chemostat in steady state may either contain no biomass (black region), only strain P (light grey), only strain C (dark grey), or both P and C (color). Colors from blue to yellow (see color bar) indicate the ratio of C's glucose consumption rate and P's glucose consumption rate when both strains are present in the chemostat, which varies between zero (when C does not consume glucose) and one (when C and P consume glucose at the same rate). The glucose consumption ratio is large (green, orange and yellow) when the producer strain produces little acetate (small $p_{ac,P}$) while the consumer strain C consumes a lot of glucose (high $c_{glc,C}$). The maximally possible glucose consumption ratio of one (bright yellow) is observed only when both strains are metabolically indistinguishable, i.e., when P produces no acetate and C uses glucose as the only source of carbon. The lower panel magnifies part of the upper panel for low values of acetate production ($p_{ac,P} < 30$) to appreciate how glucose consumption changes at these values.

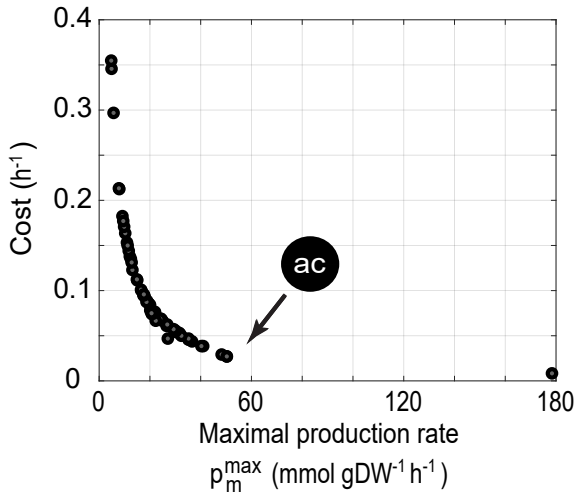


Figure 2.7: Correlation between maximal production rate and cost. The figure shows the maximal production rate and cost for each of 58 secondary carbon source (grey circles) that *E. coli* can produce when growing on glucose. The black arrow indicates the data for acetate. See supplementary 2.6.2 Text for a description of how these quantities are calculated.

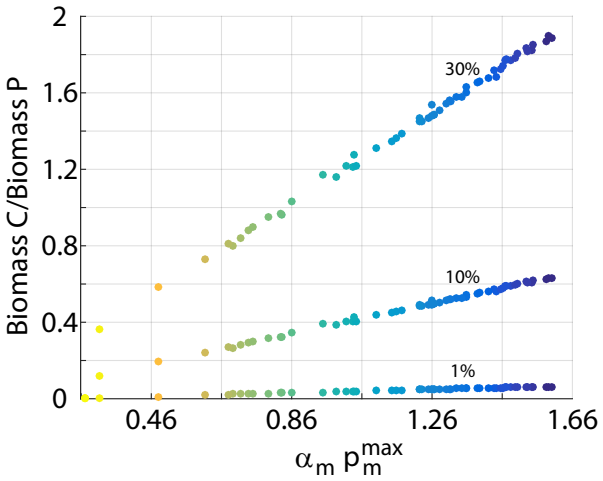


Figure 2.8: Steady state biomass ratios for different secondary carbon sources. The figure is a different representation of the data shown in Figure 2.2D. The vertical axis shows the steady state biomass ratio C/P as a function of the product (horizontal axis) of maximal production (p_m^{max}) and biomass yield (α_m) of the secondary carbon sources considered here (horizontal axis). Each circle in the plot corresponds to a secondary carbon source that *E. coli* can produce when glucose is the primary carbon source. Circle colors indicate the product of maximal production (p_m^{max}) and biomass yield (α_m) of a secondary carbon source. This product equals 1.26 h^{-1} for acetate. The biomass ratio is shown for 1, 10 and 30% of the maximally possible secondary carbon source production flux, as indicated by the numbers in the panel.

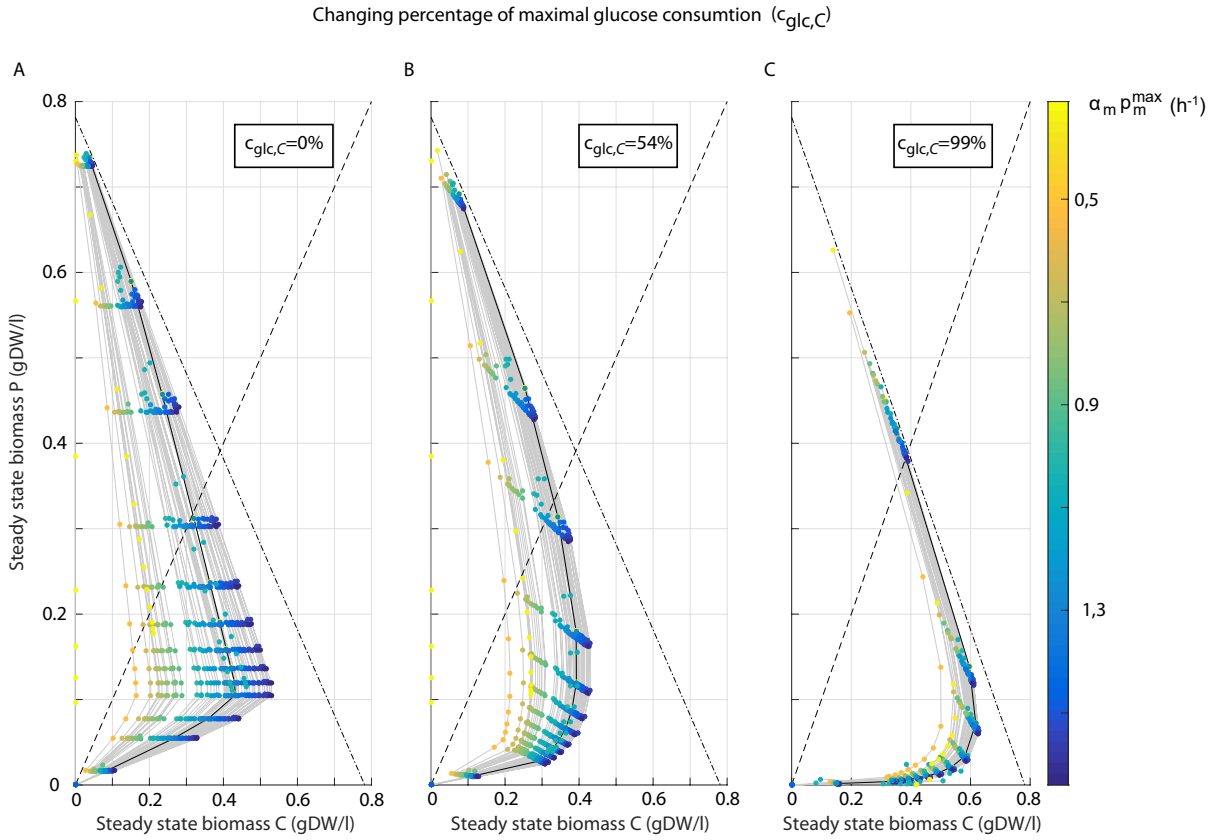


Figure 2.9: Steady state biomass of all cross-feeding strain pairs, when glucose is the primary carbon sources and when the consumer strain consumes different amounts of glucose. Steady state biomass of producer strain P (vertical axes) and consumer strain C (horizontal axes) at different percentages of the maximum synthesis rate (p_m^{max}), i.e., the synthesis rate of the secondary carbon source beyond which the producer strain P is flushed out of the chemostat, for each of 54 secondary carbon sources (grey lines, one line per carbon source). Circles are placed at 1, 5, 10-90, 95 and 99% of the maximum synthesis rate. The dashed-dotted line indicates the maximally achievable total steady-state biomass (0.78 gDW/l), which is obtained when glucose is metabolized completely to CO2 by the producer strain, without synthesis of any secondary carbon source. The dashed line indicates where both strains have identical biomass. Circle colors indicate the product of maximal production (p_m^{max}) and biomass yield (αm) of a secondary carbon source. This product equals 1.26 h^{-1} for acetate (black line superposed with blue circles).The consumer strain has a consumption rate of glucose $c_{glc,C}$ that is equal to 0% (A), 54% (B), and 99% (C) of the rate it needs to persist on glucose alone in the chemostat. Panel A is identical to Figure 2.2D and merely shown to facilitate comparison.

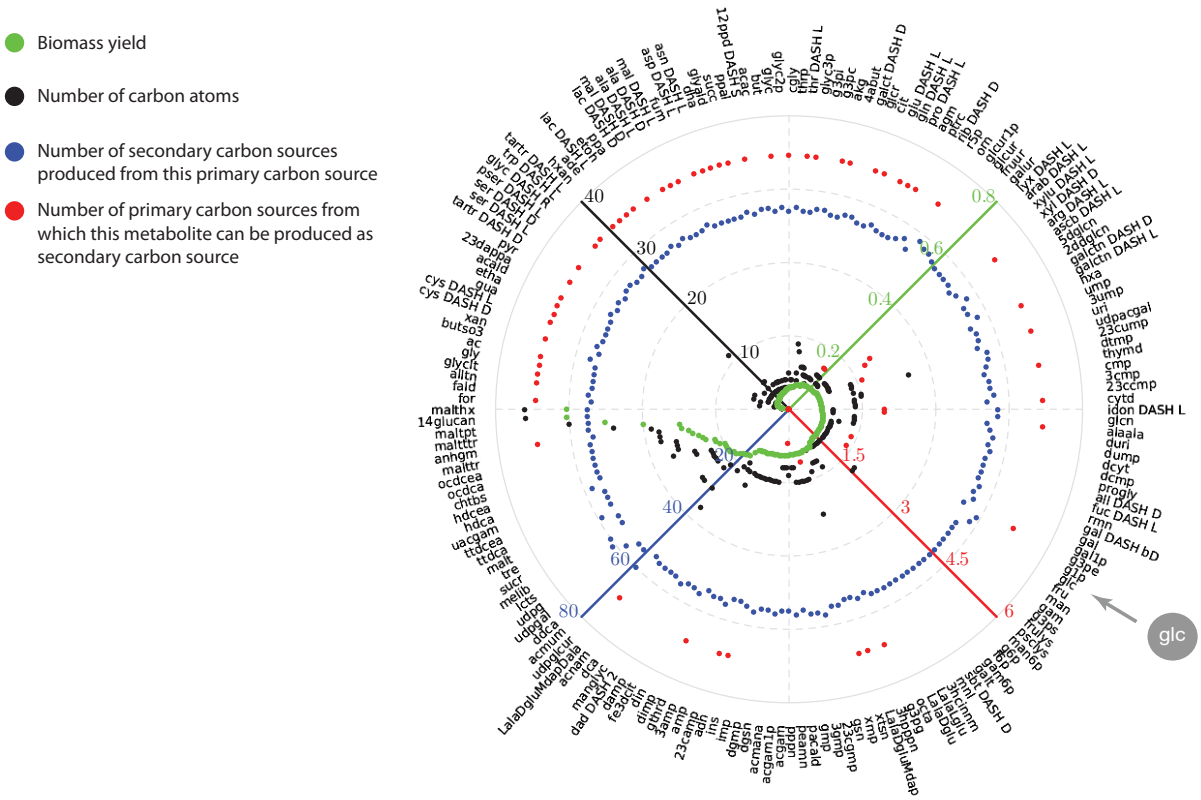


Figure 2.10: Carbon sources and associated statistics for *E. coli*. The outermost circle lists all of *E. coli*'s carbon sources (as in Figure 2.2A), ordered clockwise according to biomass yield, starting from formate (for, 9' o'clock). Green circles (solid green scale bar from center to top right) indicate the biomass yield of each carbon source. Black circles (solid black scale bar from center to top left) indicate the number of carbon atoms of the carbon source. Blue circles (solid blue scale bar from center to lower left) indicate the number of secondary carbon sources that can be produced when *E. coli* grows on a given primary carbon source. Red circles (solid red logarithmic scale bar from center to down right) indicate from how many primary carbon sources this carbon source can be produced as a secondary carbon source (if it can be produced at all). Most secondary carbon sources can be produced from all primary carbon sources (ln(179)=5.2 on the red scale), but some can be produced only from few primary carbon sources. Among them is glucose (grey arrow), which can be produced only from four primary carbon sources.

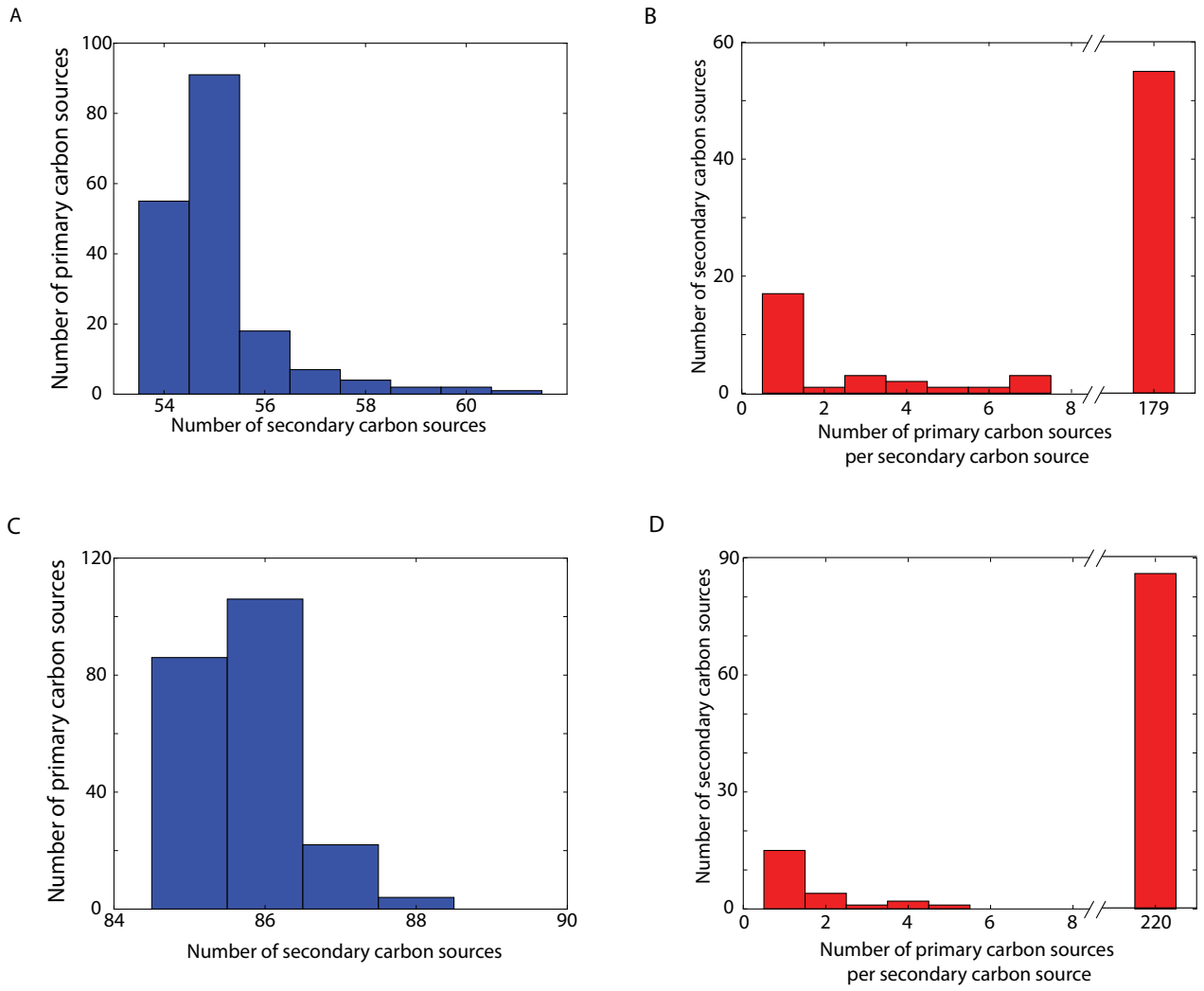


Figure 2.11: Primary and secondary carbon sources in *E. coli* and in the pan-metabolic network. (A) Histogram of the number of secondary carbon sources that can be produced per primary carbon source in *E. coli* (blue dots in supplementary 2.10 Fig). (B) Histogram of the number of primary carbon sources from which each of the 83 secondary carbon sources in *E. coli* can be produced (see also red circles in supplementary 2.10 Fig). (C) and (D), like (A) and (B) but for the pan-metabolic network.

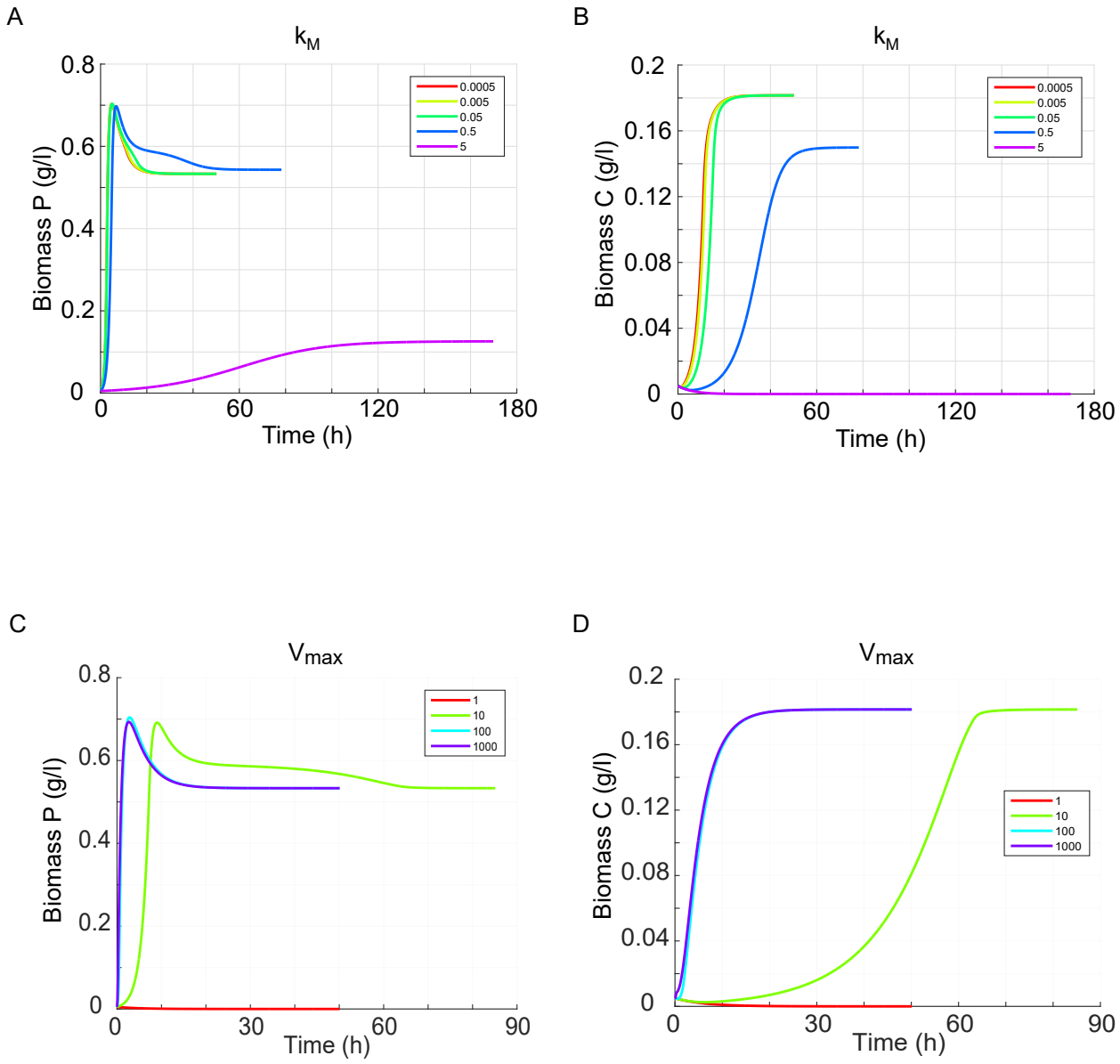


Figure 2.12: Chemostat dynamics of a glucose consumer-acetate producer community for different transport-associated parameters. We model transport limitation with Michaelis-Menten kinetics with parameters V_{max} and k_M (Methods). Horizontal axes in all panels indicate time in hours. Panels (A) and (B) show the biomass values of P and C, respectively (vertical axes), as a function of time, while changing k_M (expressed in mM, see color legends in both panels). The range of k_M values simulated covers 60% of all k_M values present in Brenda database (Pharkya et al., 2003) (The median k_{Min} the database is approximately 0.1 mM, with 60% of k_M values between 0.001 and 1 mM (Bar-Even et al., 2011)).Panels (C) and (D) show the biomass values of P and C, respectively (vertical axes), as a function of time, while changing V_{max} (in $mmol\, gDW^{-1}\, h^{-1}$,see color legends in both panels). If, for any one consumed metabolite, V_{max} is not high enough to permit growth at the dilution rate, the population will go extinct. Values of V_{max} above this minimalvalue that permits growth at the dilution rate can alter the transient biomass dynamics but does not affect the steady state biomass. The chemostat composition and all other parameters used in these simulations are identical to those described in the main text (Section "Simulating chemostat dynamics with dynamic FBA" in methods). We ended any one simulation when a population had reached steady state, which is the reason why the lines end at different time points.

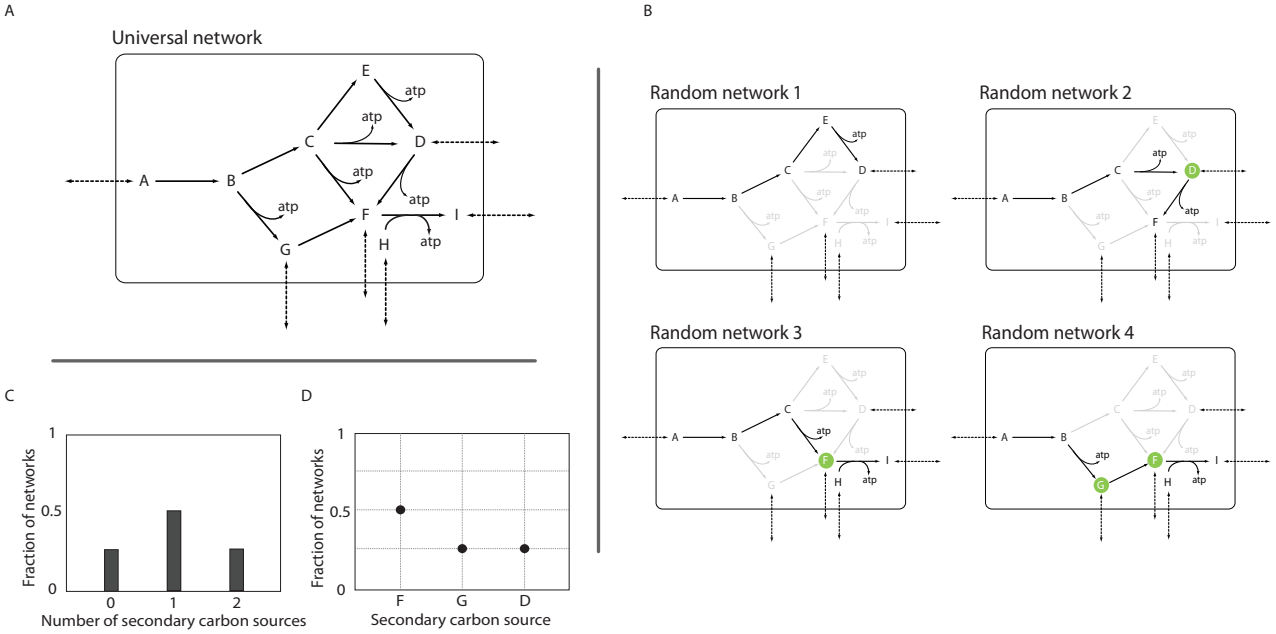


Figure 2.13: Hypothetical example to illustrate the concepts of primary and secondary carbon source and the relationship between pan-metabolic network and random viable networks. Panel (A) shows a hypothetical pan-metabolic network comprising 10 internal and 5 transport reactions. Panel (B) shows four networks created by randomly selecting 4 internal reactions (in black) from the pan-metabolic network. In this hypothetical example, ATP must be produced from at least one of the environmental nutrients (A, D, F and G) for a network to be viable. All four networks are viable on A if nutrient H is available in the environment. The random networks 2, 3 and 4 are also viable on metabolites D, F and F and G respectively. If a metabolite on which a network is viable can be produced when nutrient A is consumed, the metabolite is a secondary carbon source (green). (C) Histogram of the number of secondary carbon sources per random network, for the four random networks considered here. (D) Rank plot of secondary carbon sources (horizontal axis) that can be produced by at least one random network when A is used as a primary carbon source, ranked by the fraction of random viable networks (vertical axis) by which the secondary carbon source can be produced.

Chapter 3

The possible and the actual

Submitted as:

San Roman M. and Wagner A. (2020) Acetate and glycerol are not uniquely suited for the evolution of cross-feeding in *E. coli*

3.1 Summary

The evolution of cross-feeding among individuals of the same species can help generate genetic and phenotypic diversity even in completely homogeneous environments. Cross-feeding *Escherichia coli* strains, where one strain feeds on a carbon source excreted by another strain, rapidly emerge during experimental evolution in a chemically minimal environment containing glucose as the sole carbon source. Genome-scale metabolic modeling predicts that cross-feeding of 58 carbon sources can emerge in the same environment, but only cross-feeding of acetate and glycerol has been experimentally observed. Here we use metabolic modeling to ask whether acetate and glycerol cross-feeding are especially likely to evolve, perhaps because they require less metabolic change, and thus perhaps also less genetic change than other cross-feeding interactions. However, this is not the case. The minimally required metabolic changes required for acetate and glycerol cross feeding affect dozens of chemical reactions, multiple biochemical pathways, as well as multiple operons or regulons. The complexity of these changes is consistent with experimental observations, where cross-feeding strains harbor multiple mutations. The required metabolic changes are also no less complex than those observed for multiple other of the 56 cross feeding interactions we study. We discuss possible reasons why only two cross-feeding interactions have been discovered during experimental evolution, and argue that multiple new cross-feeding interactions may await discovery.

3.2 Introduction

One trillion microbial species have been predicted to inhabit our planet (Locey and Lennon, 2016). To understand how life on earth became so enormously diverse is a central goal of ecology and evolutionary biology. For many decades, most biological diversity was thought to arise when populations become physically subdivided (Coyne, 1992), allowing mutations to accumulate independently in each subpopulation. More recently, biologists have increasingly accepted that populations can also diversify without any physical barrier (Kondrashov and Kondrashov, 1999; Dieckmann and Doebeli, 1999b) when organisms specialize and adapt to different niches available in a heterogeneous environment (Smith, 1966; Kassen, 2002). For instance, when apples were introduced to North America, some apple maggot flies changed their plant host from hawthorn to apple. Today, apple maggot fly populations feed on hawthorns or apples. Such emerging ecological barriers can lead to the creation of new species.

Remarkably, diversity can also evolve in homogeneous environments (McKinnon and Rundle, 2002; Barluenga et al., 2006; Helling et al., 1987; Rozen and Lenski, 2000a; Good et al., 2017). Perhaps the most striking example involves stable genetic polymorphisms that originated in populations of *Escherichia coli* cultured in homogeneous batch or chemostat environments. Initially isogenic populations of an ancestral *E. coli* strain developed genetic polymorphisms which coexisted over hundreds

of generations in eleven out of fifteen evolution experiments performed in a chemostat (Helling et al., 1987)–a culturing device in which a cell culture is kept in a constant environment with the continuous addition of fresh medium and the removal of culture liquid containing leftover nutrients, metabolic waste products, and micro-organisms. The chemical environment used in these experiments was a minimal medium containing glucose as the only carbon source. Diverse strains isolated after approximately 800 generations from one of these parallel experiments showed that glucose-acetate and glucose-glycerol cross-feeding enabled the coexistence of these strains. That is, one strain consumed the primary carbon source present in the medium (glucose) and excreted a secondary carbon source (acetate or glycerol), whereas the other strains fed on the excreted secondary carbon source. Genome sequencing of the ancestral and evolved strains (Kinnersley et al., 2014) revealed almost 600 mutations in the evolved strains. Approximately 30 repeatedly mutated genes encode enzymes involved in glucose uptake, central metabolism, fermentative pathways, the TCA cycle, glyoxylate shunt and phospholipid biosynthesis. The ancestral strain itself harbored further regulatory mutations in genes required for acetate and glycerol catabolism, which may have predisposed it to evolve acetate and glycerol cross-feeding interactions (Kinnersley et al., 2009, 2014). The polymorphisms that evolved in the other parallel experiments have not been so thoroughly analyzed, but it has been shown that acetate cross-feeding was also responsible for the maintenance of five other polymorphism.

Similar cross-feeding emerged when the experiment was performed in batch culture, an environment different from a chemostat, where nutrients get depleted, waste products accumulate, and cell densities rise over time, before a sample of the culture is transferred into fresh medium. In this experiment, an isogenic population of *E. coli* cultured in a minimal glucose medium also diversified into coexisting strains which persisted for at least 10000 generations in nine out of twelve populations (Good et al., 2017). Genome sequencing of two such strains in one of these populations showed that they emerged after 6500 generations and coexisted due to glucose-acetate cross-feeding (Rozen and Lenski, 2000a). Their emergence was possibly facilitated by the population's hypermutator phenotype – the clones harbored on average 199 mutations (Plucain et al., 2014).

Experiments like these suggest that *E. coli* readily diversifies genetically and metabolically in a completely homogeneous environment by filling niches which do not exist in the environment but are created by the organism itself. We are interested in finding out how much microbial diversity can be created through such cross-feeding by characterizing the whole spectrum of molecules (beyond acetate and glycerol) that can be cross-fed. In recent work we discovered through metabolic modeling that all metabolic systems have a large potential for the evolution of cross-feeding interactions. For example, we found that when *E. coli* feeds on glucose, 58 metabolites can be excreted as by-products of metabolism. Each of these metabolites can in turn serve as a carbon source that can help sustain a stable community of cross-feeding strains. Among these metabolites are acetate and glycerol, for

which cross-feeding was observed experimentally. In other words, metabolic modeling predicts 56 additional cross-feeding interactions, which raises the question why none of these interactions have been observed experimentally. Possibly, many other such polymorphisms have indeed evolved but went undetected, because currently no systematic experimental screen for cross-feeding interactions exists. (The cross-feeding strains were detected through colony morphologies on agar plates, and substantial biochemical and genetic work was needed to prove that their polymorphisms resulted from cross-feeding.) A second possible reason is that not all predicted cross-feeding polymorphisms can evolve with the same likelihood. For example, metabolites may differ in the number of metabolic changes or DNA mutations needed to turn a strain into a producer or consumer of the metabolite. Here we explore this second possibility. That is, we ask whether glucose-acetate and glucose-glycerol cross-feeding have been observed because they are much more likely to evolve than other cross-feeding interactions. Ideally, to quantify the likelihood that a given cross-feeding interactions evolves, it would be necessary to know all mutations that give rise to the evolution of a producer or consumer strain, the probability that each mutation takes place, and the mutation's fitness effect in every genetic background and in the population in which it occurs. This amount of information is not within reach of current technology, especially because the genetic changes leading to cross-feeding may be complex and involve changes in metabolic enzymes, regulatory molecules, and transport proteins (Kinnersley et al., 2014; Plucain et al., 2014). The problem is aggravated by the fact that the same phenotypic change, such as the emergence of cross-feeding, can often be achieved through multiple and perhaps myriad different combinations of genotypic changes (Tenaillon et al., 2012; Hong and Gresham, 2014; Kryazhimskiy et al., 2014).

Faced with these obstacles, we here take a phenomenological approach, in which we estimate the likelihood for a cross-feeding interaction to emerge through the amount of metabolic change that an ancestral strain must experience to bring forth a producer and consumer strain for the cross-fed metabolite. In other words, we use an assumption of parsimony: The producer and consumer strains most likely to evolve from an ancestor are those that are metabolically most similar to the ancestor.

More specifically, we perform the following analysis for each of 58 metabolites that can form the basis of a stable cross-feeding polymorphism emerging in a chemostat initialized with an ancestral *E. coli* strain and supplied with a glucose-minimal medium. We analyze the genome scale metabolic model of *E. coli* iJO1366 to identify the distribution of metabolic fluxes – the rates at which enzymatic reactions proceed – that are most similar between a producer strain of the metabolite and the ancestor, as well as between the consumer strain of this metabolite and the ancestor. We use multiple measures of similarity, among them the number of reactions that require a change in flux, and we assume that a producer or consumer strain is more likely to emerge if this number is small. Our analysis shows that acetate and glycerol cross-feeding do not require exceptionally small metabolic changes compared to

the 56 other metabolites we consider.

3.3 Results

3.3.1 Complex metabolic changes are needed for the evolution of glucose-acetate and glucose-glycerol cross-feeding.

The four stably co-existing and cross-feeding *E. coli* strains that evolved from an ancestral strain in a glucose-limited chemostat showed genetic and physiological differences with respect to glucose, acetate, and glycerol uptake and metabolism (Helling et al., 1987). Our first analysis prepares the ground by modeling the ancestral and evolved flux distributions of *E. coli* that are relevant to reconstruct the lab-evolved glucose-acetate and glucose-glycerol cross-feeding interactions. We began by modeling the metabolic behavior of the ancestral strain using a modified version of Flux Balance Analysis (FBA) known as parsimonious FBA (pFBA), together with a genome scale metabolic network of *E. coli* (iJO1366) (see Methods).

Flux balance analysis is a computational method for predicting metabolic fluxes (the rates at which substrates are converted into products) for all reactions in a genome-scale metabolic network. Essentially, FBA identifies a flux distribution that results in maximal biomass growth while fulfilling a set of constraints. These include the assumption that metabolism operates in a steady-state, where the production and consumption of each metabolite are exactly balanced. The constraints also include assumptions about the reversibility of reactions, as well as about maximally possible rates of nutrient transport. Generally there are multiple alternative flux distributions that fulfill all constrains and permit maximal growth. Among them, parsimonious FBA (Lewis et al., 2010) identifies the flux distribution that minimizes the sum of all metabolic fluxes, which can be viewed as a proxy for the total expression level of metabolic enzymes. In other words, pFBA assumes that a metabolism must achieve maximal growth subject to minimal cost. This cost minimization increases consistency between computational predictions and transcriptomic and proteomic data (Lewis et al., 2010).

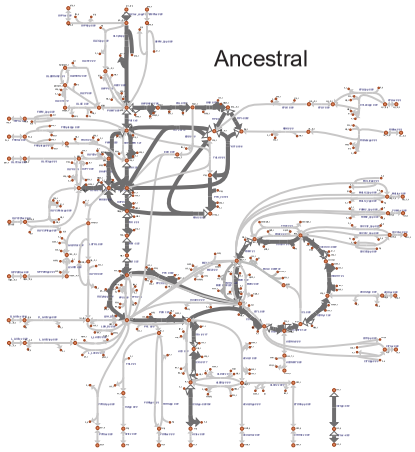
We applied pFBA to determine the flux distribution of the ancestral strain from which cross-feeding emerged under the conditions in which its evolution had been observed experimentally (Helling et al., 1987). Specifically we assumed a chemically minimal environment where sufficient glucose – the sole carbon source – is available to allow growth at the dilution rate of the chemostat in the experiments that inspired this work ($0.2\ h^{-1}$) (Helling et al., 1987). (See methods for details.) Figure 3.1B illustrates part of the resulting flux distribution graphically for central carbon metabolism. After having obtained this ancestral flux distribution we predicted the flux distributions of the evolved cross-feeding strains. We used the same metabolic network of *E. coli* (iJO1366) to model ancestral and cross-feeding strain. This modeling decision reflects the observation that cross-feeding strains can emerge in little evolutionary

time (Helling et al., 1987; Rozen and Lenski, 2000a), and that metabolic differences between strains are not due to differences in their enzyme-coding genes, but result from mutations that affect the expression of these genes or the activity of the encoded enzymes. In addition, we assumed that evolution is most likely to bring forth producer and consumer strains whose flux distribution differs as little as possible from the ancestor.

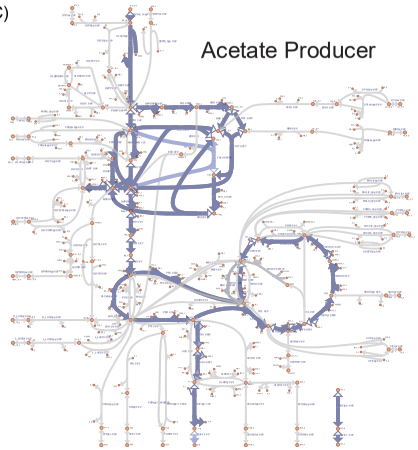
(A)

	Acetate		Glycerol	
	Producer	Consumer	Producer	Consumer
N. of reactions with a change in flux (Distance to ancestor)	41	60	45	43
Metabolic subsystem				
Glycolysis/Gluconeogenesis	7	11 on/off	9 on/off	9 on/off
Pentose phosphate pathway	8 on	10 off	10 on	8 on/off
Citric Acid Cycle	8	8	8	8
Transport inner/outer membrane	10 on	10 on/off	10 on	9 on/off
Others	8	21 on/off	8 on/off	9 on
N. of genes	90	119	93	92
N. of operons	52	78	56	53
N. of regulons	35	39	36	36

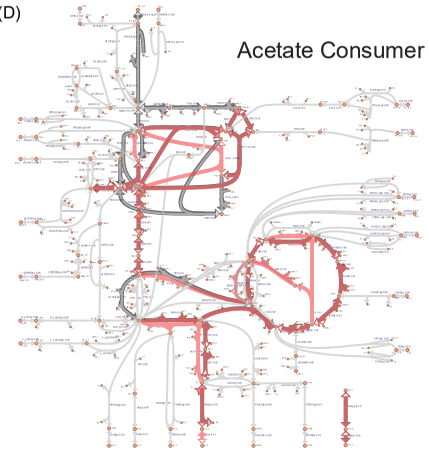
(B)



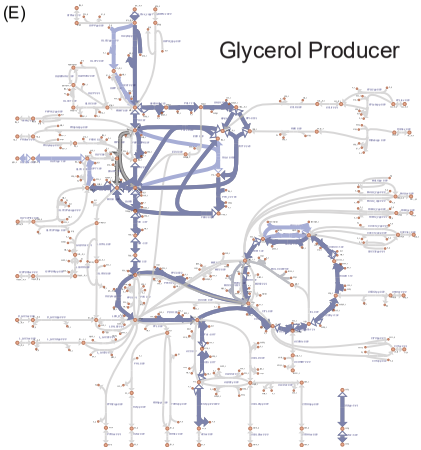
(C)



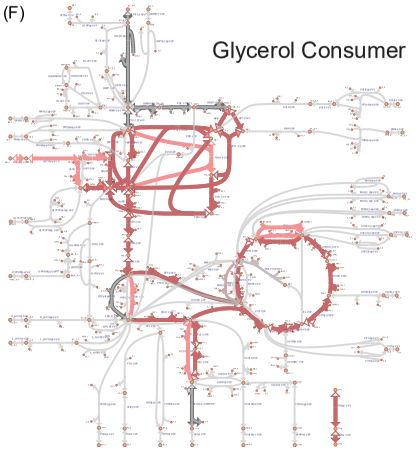
(D)



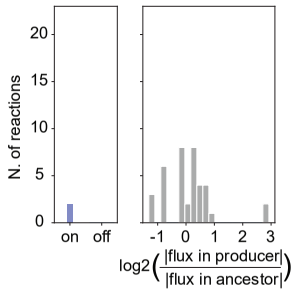
(E)



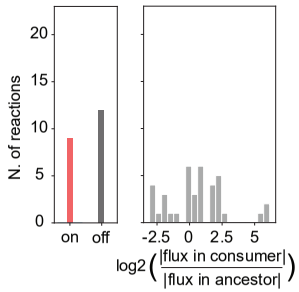
(F)



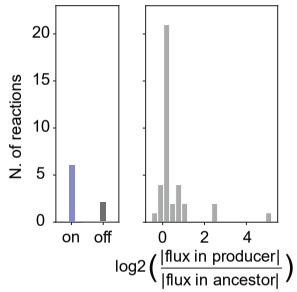
(G)



(H)



(I)



(J)

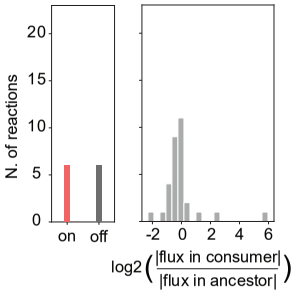


Figure 3.1 ***(previous page)*****: Metabolic changes required for the evolution of the acetate and glycerol cross-feeding strains.** (A) Minimal number of reactions requiring a change in flux in the ancestor for acetate and glycerol cross-feeding to evolve. Most reactions requiring a flux change belong to glycolysis and gluconeogenesis, the pentose phosphate pathway, the citric acid cycle or from transport processes. The remaining reactions ('others') come from multiple pathways that comprise oxidative phosphorylation, alternate carbon metabolism, pyruvate metabolism, glycine and serine metabolism, alanine and aspartate metabolism, folate metabolism, anaplerotic reactions, or that are unassigned to a pathway. In the table we also include the number of genes associated with the reactions requiring a flux change, the number of operons into which these genes fall, as well as the number of regulons. (B) Central carbon metabolism of *E. coli*. Every orange circle represents a metabolite and every line a reaction. Thick grey lines indicate a non-zero flux in the ancestral strain ($|\alpha_i|$>0.001 $mmol\, gDW^{-1}\, h^{-1}$). (C) to (F) As in (B), but each panel shows reactions with non-zero flux in cross-feeding strains (blue for producers, red for consumers) in addition to non-zero fluxes in the ancestor (grey). Panels (G) to (J) show, on the left of each panel, the number of reactions that are activated in a producer (blue) or consumer (red) relative to the ancestor ('on'), or the reactions that are inactivated relative to the ancestor ('off', grey). On the right of each panel, the amount of flux change is shown for reactions that change their flux relative to the ancestor.

We first focused on the origin of the strain that produces acetate. Experimentally, this strain was found to consume more glucose than the ancestor, and to excrete acetate and glycerol into the environment. To disentangle the metabolic changes that are required for acetate and glycerol production, we modeled two separate producer strains: an acetate producer and a glycerol producer. To model the acetate producer strain, we imposed non-zero ($1\ mmol\,gDW^{-1}\,h^{-1}$) acetate excretion on this strain and allowed the strain to consume more glucose than the ancestral strain, because it could otherwise not persist in the chemostat. Specifically, we set glucose consumption to the minimal amount required to satisfy the acetate excretion constraint and to allow growth at the dilution rate value. To reflect our parsimony (minimal metabolic change) assumption, we assumed that acetate production in this strain is achieved via the smallest possible flux rearrangement relative to the ancestral strain. More specifically, we used Regulatory On/Off Minimization (ROOM), an optimization method that finds the flux distribution which minimizes the number of reactions whose flux needs to change from the ancestral distribution such that the strain can produce acetate.

We modeled the evolution of the second cross-feeding strain, the acetate consumer, analogously. Experimentally, this strain was found to consume glucose but distinguished itself from the ancestor and the other evolved strains by its large acetate consumption capacity. We modeled the acetate consumer strain by disallowing glucose consumption completely and permitting only acetate consumption, because it allows us to identify the metabolic changes that are associated with a change in carbon source most clearly. (We also repeated the analysis allowing the acetate consumer strain to consume both acetate and glucose, which led to the same conclusions. See supplementary 3.4 Fig.)

Again, we applied ROOM to identify the smallest number of reactions whose flux needs to change to bring forth an acetate consumer strain. We then repeated this entire procedure for both the glycerol producer and consumer. Figures 3.1C-F show the flux through central carbon metabolism in the producers (in blue) and consumers (in red) on top of the flux distribution identified for the ancestral strain (in grey). We found that acetate (glycerol) production and consumption requires changes in fluxes through at least 41 (45) and 60 (43) reactions (Figure 3.1A). In other words, the required flux change, even though it is the minimally necessary change, is complex.

This complexity is also evident in different kinds of predicted flux changes. Some reactions are active (non-zero flux) in the ancestor but inactive ('off') in the evolved strain. Other reactions are active ('on') only in the evolved strain. Yet other reactions change only their flux magnitude in the evolved strain. Figure 3.1G-J show the numbers of reactions in these three categories. Although the majority of reactions (65-95%) change only their flux magnitude, between 2 and 21 reactions need to be turned on or off to allow the production or consumption of acetate and glycerol.

In addition to comprising different kinds of changes, the changed fluxes fall into various metabolic subsystems, including glycolysis and gluconeogenesis, the pentose phosphate pathway, the citric acid

cycle, and transport (Figure 3.1A). Thus, it is unlikely that they could be brought forth by one or few mutations, an assertion that is corroborated by our next analysis.

To find out how flux changes might be related to genetic changes, we used the Gene-Protein-Reaction association (GPR) map available for the metabolic model of *E. coli* iJO1366. The GPR map is only one-to-one in the simplest case, where a gene product catalyzes one reaction. Alternatively, a gene product may catalyze more than one reaction; the products of multiple genes may be needed to catalyze one reaction; or the products of different genes may catalyze the same reaction. The 41 reactions requiring a flux change to bring forth acetate producer strain are linked to 90 genes, which are organized in 52 operons and 35 regulons. Figure 3.1A shows the number of genes, operons and regulons associated with the reactions requiring a flux change for the evolution of the acetate consumer and glycerol cross-feeding strains. The changes involve multiple operons and regulons. Consistent with the prediction that one or few mutations could not bring about all these changes, experimentally evolved cross-feeding strains harbored hundreds of mutations. Multiple repeatedly mutated genes (Kinnersley et al., 2014) were directly involved in glycolysis and gluconeogenesis, the TCA cycle, and transport, which are three of the subsystems where we also observe most of the reactions changes. Transcriptomic changes (Kinnersley et al., 2014, 2009) showed very limited agreement with computational predictions (see supplementary 3.6.1 text and supplementary 3.5 Fig). This is unsurprising, because gene expression change poorly reflects metabolic flux change for several reasons, for example because mRNA and enzyme abundance correlate poorly (Futcher et al., 1999; Greenbaum et al., 2003b; Ideker et al., 2001; Washburn et al., 2003; Vogel and Marcotte, 2012; Guimaraes et al., 2014).

3.3.2 Acetate and glycerol cross-feeding does not require exceptionally little metabolic change.

A previous analysis of the metabolic network of *E. coli* showed that 56 distinct cross-feeding interactions other than the experimentally described glucose-acetate and glucose-glycerol interactions can evolve and lead to stable polymorphic communities (San Roman and Wagner, 2018). This raises the question why only the latter two types of cross-feeding polymorphisms have been described. One possible answer is that the evolution of acetate and/or glycerol cross-feeding requires less metabolic changes than cross-feeding involving other metabolites, and is thus more likely to evolve. To find out, we repeated the analysis from the previous section, but replaced the secondary carbon sources acetate and glycerol with each of the other 56 metabolites that could potentially be cross-fed.

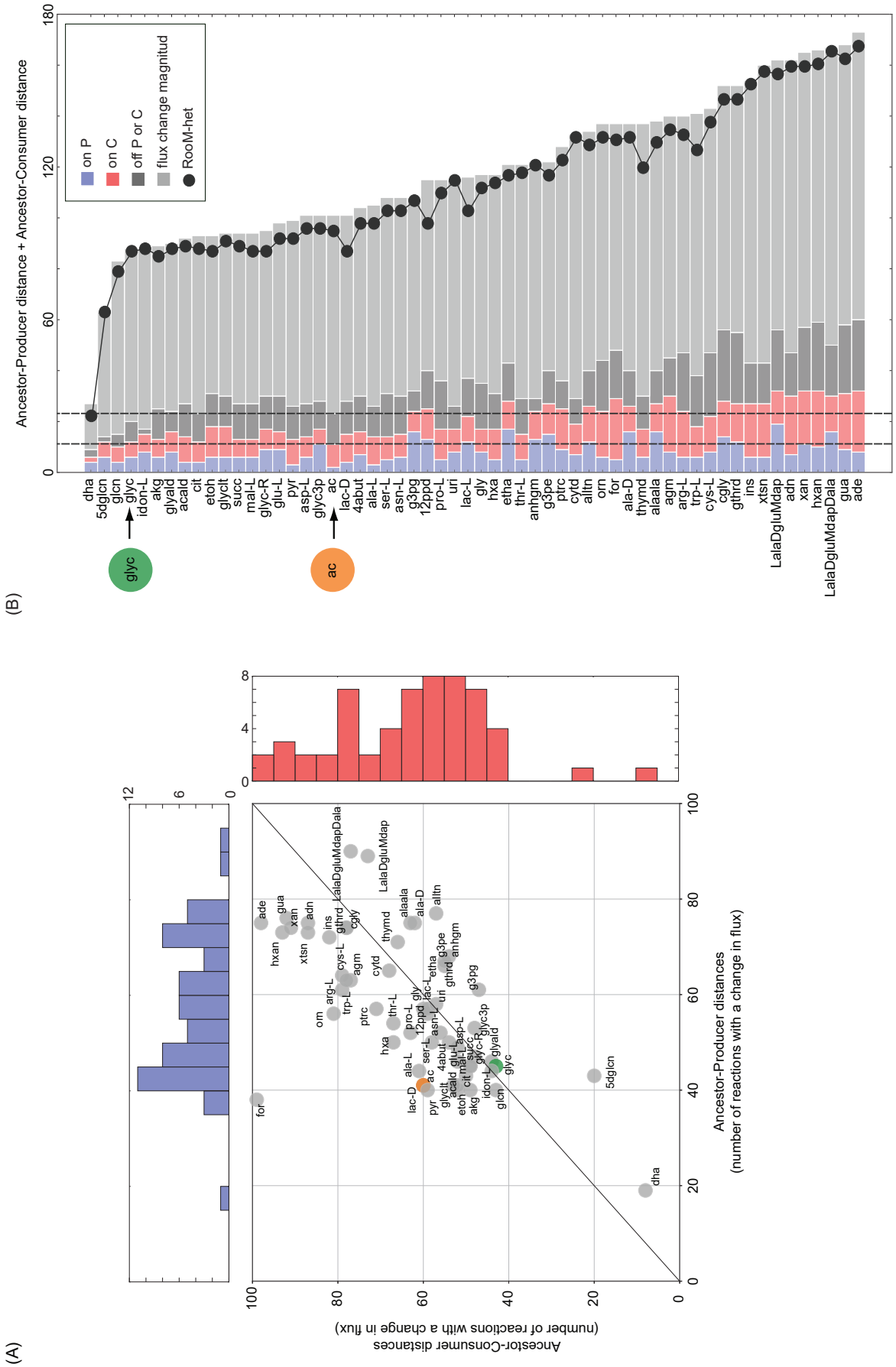

(A)
Ancestor-Consumer distances
(number of reactions with a change in flux)
Ancestor-Producer distances
(number of reactions with a change in flux)
(B)
Ancestor-Producer distance + Ancestor-Consumer distance
on P
on C
off P or C
flux change magnitud
RooM-het
glyc
ac

Figure 3.2 ***(previous page)*****: Metabolic distance between the ancestor and cross-feeding strains.** (A) The ancestor-producer and ancestor-consumer distances (measured as the number of reactions requiring a flux change) obtained when performing RooM are given on the x- and y- axes respectively. Every circle in the plot corresponds to one metabolite with cross-feeding potential. Acetate and glycerol are shown as orange and green circles respectively. The diagonal line is shown as a visual guide. A circle on the line indicates that the same number of reactions needs to change their flux to create the corresponding producer and consumer strain. Blue and red histograms show the distribution of the ancestor-producer and ancestor-consumer distances, respectively. (B) The y-axis shows acronyms for the 58 metabolites that can lead to stable cross-feeding interactions, ranked according to decreasing probability of evolving cross-feeding, as quantified by the total RooM-predicted distance (sum of ancestor-producer and ancestor-consumer distances) shown on the x-axis. Different bar colors indicate the number of reactions classified as turned 'on' in the producer relative to the ancestor (blue), turned 'on' in the consumer relative to the ancestor (red), turned 'off' in either the producer or consumer relative to the ancestor (dark grey), as well as reactions requiring a flux change in either producer or consumer relative to the ancestor ('flux change magnitude', in light grey). Black circles show the total distances obtained when RooM-het was used to minimize strain distances.

The minimal metabolic distances between ancestor and producer, as well as between ancestor and consumer, as obtained with RooM, are shown in figure 3.2A. The mean producer-ancestor distance equaled 57±15 reactions with changed flux. The corresponding consumer-ancestor distance equaled 62±17 reactions. In 41 out of the 58 cross-fed metabolites (those situated above the diagonal in figure 3.2A), the producer of a given metabolite can evolve more easily than its consumer, requiring fewer flux changes.

The cross-feeding of seven carbon sources requires less flux change to create both producer and consumer strains than acetate cross-feeding. These carbon sources are dihydroxyacetone (dha), α-ketoglutarate (akg), pyruvate (pyr), glycolate (glyclt), ethanol (etoh), acetaldehyde (acald) and D-gluconate (glcn). Likewise, cross-feeding on dihydroxyacetone (dha), D-gluconate (glcn) and 5-Dehydro-D-gluconate (5dglcn) requires less flux change than cross-feeding on glycerol.

Figure 3.2B shows 58 metabolites ranked by the likelihood that cross-feeding evolves for them. Based on our minimal change criterion, cross-feeding of 18 metabolites is easier to evolve – it requires fewer reaction changes in producer and consumer. In contrast, cross-feeding of only three carbon sources is easier to evolve than glycerol cross-feeding. In sum, acetate and glycerol are not exceptional in their potential to evolve in cross-feeding.

Thus far, we have used the sum of the ancestor-producer and ancestor-consumer distances as a proxy of the likelihood of the cross-feeding interaction to evolve. By doing so we are inherently assuming that producers and consumers evolve independently from each other. This is strictly not correct, because the producer needs to evolve before the consumer does, otherwise the consumer will lack a carbon source on which to feed. However, because a more complex model (supplementary 3.6.3 Text) yields essentially identical predictions (see supplementary 3.7 Fig), we will continue to use the sum of metabolic distances below.

Figure 3.2B uses bars with different colors to distinguish reactions that are activated ('turned on'), inactivated ('turned off'), and that change flux magnitude relative to the ancestor. Just as for acetate and glycerol cross-feeding (figure 3.1G-J), most reactions change their flux quantitatively rather than qualitatively (being turned on or off). Of special interest are those reactions that change flux qualitatively, because it is possible that such flux change is more difficult to achieve genetically, for example because fewer mutations eliminate a gene than modulate its activity (Eyre-Walker and Keightley, 2007). For acetate cross-feeding to emerge, 23 reactions are turned on or off, a number that is lower for five other carbon sources. Likewise, for glycerol-cross feeding to emerge, 20 reactions are turned on or off, a number that is lower in four other carbon sources. Thus, even if the number of reactions turned on or off were the most appropriate measure of metabolic distance, acetate and glycerol cross-feeding would not be exceptional in their metabolic distance to the ancestor. Cross-feeding of other carbon sources requires even fewer qualitative changes than cross-feeding of acetate

and glycerol.

The reactions which change flux overlap to some extent across different cross-fed carbon sources. Specifically, out of the 2583 reactions present in *E. coli*'s model iJO1366, 392 (390) reactions are required to change their flux to convert the ancestor into a producer (consumer) of at least one of the 58 carbon sources. Nine (four) reactions require a flux change for the evolution of every producer (consumer). 36 (40) changing reactions are shared among 80% of all producer (consumer) strains. This overlap suggests that the mutations required to create producer and consumer strains of different metabolites may also overlap.

In sum, if metabolic distance is an appropriate proxy for the likelihood that cross feeding evolves, acetate and glycerol cross-feeding are not exceptionally likely to evolve compared to cross-feeding on 56 other carbon sources. This is further supported by our observation that the reactions that require a flux change for the evolution of producers or consumers overlap among different cross-fed carbon sources.

3.3.3 Can a heterogeneous ancestral population affect the likelihood that cross-feeding emerges?

Thus far we assumed that the ancestral population is homogeneous, i.e., it is composed of phenotypically identical individuals. This is why we modeled it with a single flux distribution predicted through pFBA. However, bacterial populations are often phenotypically heterogeneous. This heterogeneity may arise from genetic differences among individuals in a population, or from noisy gene expression in genetically homogeneous (isogenic) populations. Likewise, in the chemostat experiment that inspired this work, the isogenic ancestral strain may have been phenotypically heterogeneous, or it may have diversified genetically before cross-feeding emerged. Such heterogeneity can affect ecological and evolutionary processes. In this section we ask how such heterogeneity might affect the evolution of cross-feeding. To this end, we developed a method we call "RooM-het" which allows us to identify the cross-feeding strains that most likely evolve from a heterogeneous ancestral population. As opposed to RooM, RooM-het does not use a single (ancestral) flux distribution as reference. Instead, it identifies two minimally distant flux distributions simultaneously, each fulfilling a different set of constraints (See figure 3.3A and 'Methods' for details).

We applied RooM-het to identify producer and consumer strain for each of our 58 carbon sources that can lead to cross feeding. Assuming a heterogeneous ancestral population led to a lower distance to the ancestor in producer and consumer strains. However, the distance reduction was only modest (two changed reactions on average, supplementary 3.8 Fig). As in the previous section, we used the sum of the ancestor-producer and ancestor-consumer distances as a proxy of the likelihood that the different cross-feeding interactions emerge. The black circles in figure 3.2B show the total distance

to the ancestor as obtained with RooM-het. Taking into account population heterogeneity affects the likelihood of cross-feeding only modestly, changing the rank of acetate (glycerol) cross-feeding from 19-th (4-th) to 18-th (5-th) most likely to evolve. Thus, acetate and glycerol cross-feeding are not especially likely to evolve even when heterogeneous ancestral populations are considered.

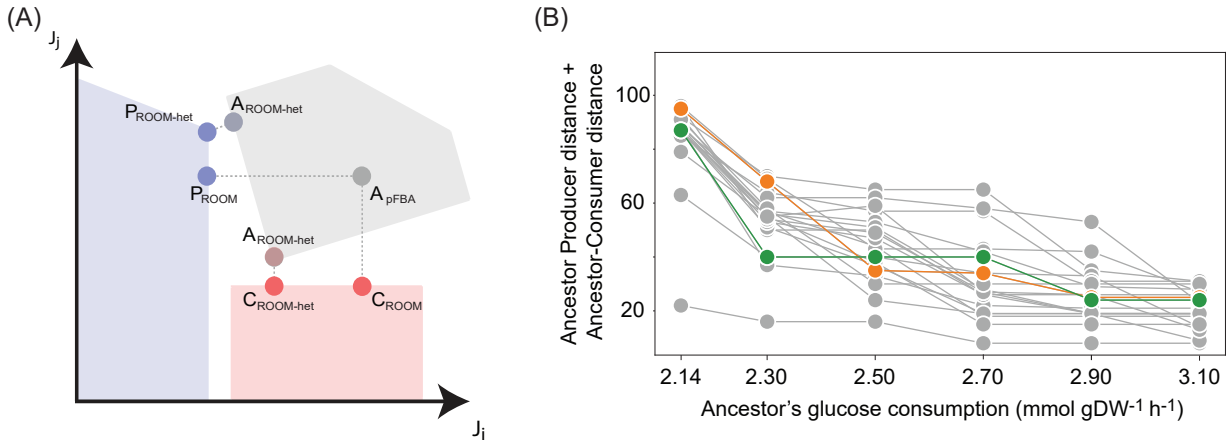


Figure 3.3: Impact of a heterogeneous ancestral population on the evolutionary outcome. (A) Explanation of the RooM and RooM-het methods. The figure shows a hypothetical flux space where the fluxes through reaction i (J_i) and j (J_j) are shown. Different carbon source consumption and production rates impose different constraints that affect the allowable solution space of ancestor, producer and consumer strains differently. The allowable solution space corresponds to the space of flux distributions that fulfill a set of constraints, here shown as grey, blue and red areas for the ancestor, producer and consumer respectively. To perform RooM we first identified an ancestral flux distribution with pFBA (here represented as the grey circle labeled A_{pFBA}). We used this flux distribution as a reference to identify the producer (blue circle labeled P_{ROOM}) and consumer (red circle labeled C_{ROOM}) flux distributions that would require the minimal number of flux changes. In contrast, RooM-het requires no ancestral reference flux distribution. When using this method to identify a producer flux distribution, the method returns two flux distributions, one that satisfies all the constraints of being a producer, and another that satisfies all the constraints of being an ancestor. The same holds when predicting a consumer flux distribution. In the figure, the resulting producer and consumer flux distributions are shown as blue and red circles labeled $P_{ROOM-het}$ and $C_{ROOM-het}$). The two distinct ancestral flux distributions that result when identifying the producer and consumer distribution are labeled $A_{ROOM-het}$. (B) Sum of the ancestor-producer and ancestor-consumer distances (vertical axis) as a function of the glucose consumed by the ancestral population (horizontal axis). Predictions for acetate and glycerol are shown in orange and green, respectively. Grey circles correspond to predictions for one of the twenty metabolites with a predicted likelihood of being subject to evolved cross-feeding greater than that observed for acetate when either RooM or RooM-het are performed at the minimal glucose consumption rate of 2.14 $mmol\,gDW^{-1}\,h^{-1}$.

3.3.4 Greater glucose consumption can modify the likelihood of cross-feeding interactions to emerge

Thus far, we assumed that the ancestral strain was maximally efficient in consuming glucose, that is, it consumed the minimal amount of glucose required for growth at a dilution rate of 0.2 h^{-1}. Reducing this efficiency, that is, allowing more than this minimal glucose consumption, may affect metabolism in multiple ways (San Román et al., 2014). It may open alternative ways of metabolizing carbon, increase the production of waste products, and in doing so, increase population heterogeneity. Here we explore how such increased glucose consumption may affect the likelihood of observing different cross-feeding interactions. We focus on 20 metabolites, which comprise acetate and the 19 metabolites (including glycerol) whose likelihood to be cross-fed is higher than acetate based on either RooM or RooM-het.

In this analysis, we employed again RooM-het but now we varied the ancestral glucose consumption from the minimum required for growth at 0.2 h^{-1} (2.14 $mmol\,gDW^{-1}\,h^{-1}$) to a value of 3.1 $mmol\,gDW^{-1}\,h^{-1}$, which corresponds to the glucose consumption required by the gluconate (glcn) producer strain, which has the highest glucose requirement. As the ancestor consumes increasing amounts of glucose, the metabolic change required for the emergence of producer and consumer strains decreases substantially (supplementary 3.9 A and B Figs). For example, when glucose consumption is minimal (2.14 $mmol\,gDW^{-1}\,h^{-1}$) the mean ancestor-producer distance for the twenty metabolites of interest equals 41±7 changed reactions, which reduces to 25±8 reactions when glucose consumption increases by only 7% (2.3 $mmol\,gDW^{-1}\,h^{-1}$). Remarkably, this ancestor-producer distance declines to zero for all carbon sources above some threshold of glucose consumption, meaning that the ancestor already produces the cross-fed metabolite without any flux change when it consumes a sufficient amount of glucose. The mean ancestor-consumer distance also decreases substantially (from 44±10 to 29±6 reactions with a 7% increase of glucose consumption), but it generally does not decline to zero. The relative proportion of reactions that are activated, inactivated, or that merely changed flux magnitude relative to the ancestor does not change as glucose consumption increases (supplementary 3.9 C to H Figs).

Given these observations, it is not surprising that the sum of the ancestor-producer and ancestor-consumer distances decreases as glucose consumption increases. However, the extent of this decrease differs between metabolites (figure 3.3B) and thus, the ranking of metabolites most likely to be cross-fed changes as glucose consumption increases. For example, Spearman's correlation coefficient of these ranks varied from a minimum of 0.42 (when changing glucose consumption from 2.3 to 2.5 $mmol\,gDW^{-1}\,h^{-1}$) to a maximum of 0.89 (when changing glucose consumption from 2.7 to 2.9 $mmol\,gDW^{-1}\,h^{-1}$). Acetate ranks highest (fifth) when glucose consumption is 2.5 $mmol\,gDW^{-1}\,h^{-1}$, in which case cross-feeding involving dihydroxyacetone (dha), D-gluconate (glcn), glycolate (glyclt)

and glyceraldehyde (glyald) are more likely to evolve. Glycerol ranks highest (fourth) at a glucose consumption of 2.3 $mmol\, gDW^{-1}\, h^{-1}$, where it is preceded by dihydroxyacetone (dha), glyceraldehyde (glyald), and 5-dehydro-D-gluconate (5dglcn).

In sum, even at elevated glucose consumption, cross-feeding of several metabolites is more likely to evolve than cross feeding of acetate and glycerol. However, this likelihood is sensitive to glucose consumption, which shows that interactions between a metabolism and its environment are critical to determine the likelihood that cross-feeding emerges.

3.4 Discussion

Based on our previous predictions that 58 different metabolites can sustain stable communities of cross-feeding *E. coli* bacteria which emerge in a glucose limited chemostat (San Roman and Wagner, 2018), we here identified the minimal amount of metabolic change (numbers of reactions with changed metabolic flux) required for the emergence of each such cross-feeding interaction. We used this amount of change as a proxy for the likelihood of cross-feeding to evolve, with more change implying a lower likelihood of evolution. Regardless of the cross-fed carbon source, the required change was complex. It involved multiple biochemical reactions, metabolic pathways, operons, and regulons. These observations are consistent with a large number of mutations and regulatory changes observed in experimentally identified cross-feeding communities (Kinnersley et al., 2014; Plucain et al., 2014).

Most importantly, our analysis predicts that the experimentally observed cross-fed metabolites acetate and glycerol are not the most likely to be involved in cross-feeding interactions. Multiple other metabolites can evolve cross feeding through similar or less metabolic change. This prediction is independent of how we quantified the amount of metabolic change (supplementary 3.6.2 Text and supplementary 3.6 Fig), how we computed to likelihood of cross-feeding to emerge (supplementary 3.6.3 Text and 3.7 Fig), or whether we took the heterogeneity of an ancestral population into account (supplementary 3.8 Fig). However, we note that the amount of metabolic change required to evolve cross feeding interactions is sensitive to the amount of glucose consumed by the ancestral population, and the extent of this sensitivity depends on the cross-fed metabolite. Thus, the likelihood to evolve cross-feeding depends not only on the reaction complement of a metabolism, but also on interactions between this metabolism and the environment.

For two reasons, cross-feeding may emerge more easily in chemostats than in batch culture. First, theory shows that for cross-feeding to evolve in batch cultures, the secondary carbon source has to be produced at a high rate (Doebeli, 2002), which reduces the likelihood that cross-feeding emerges. Second, in chemostats operating at low dilution rates (such as the ones we are considering) mutants with high affinity for the available carbon source are favored and will accumulate in the population (Rabbers et al., 2015; Wu et al., 2006; Tsen et al., 1996). Because such mutants consume a lot of

the carbon source, they may not metabolize all of it completely, and may thus excrete metabolic by-products (similarly to what occurs in overflow metabolism (Xu et al., 1999)). In other words, just as for our analysis of excess glucose consumption, producer strains can emerge with little or no metabolic flux changes, which also facilitate the emergence of consumer strains (figure 3.3B).

Our analysis has two main limitations. First, we assumed that the most frequently evolving producers and consumers are those requiring the least amount of metabolic change, which we used as a proxy for the smallest amount of genetic change. However, it is well known that the relationship between genetic and phenotypic (metabolic) change is not straightforward. Whereas some DNA mutations may affect only one biochemical reaction, others may affect multiple reactions. What is more, the same amount of phenotypic change in different individuals may be caused by different numbers or kinds of mutations (Tenaillon et al., 2012; Kryazhimskiy et al., 2014; Treves et al., 1998; Chou and Marx, 2012; Toprak et al., 2011). Although these factors will reduce any association between metabolic change and genetic change, one would expect some statistical association between the two whenever multiple mutations must be responsible for the observed phenotypic change. This is probably the case for cross-feeding, where the minimal metabolic changes affect dozens of reactions in multiple biochemical pathways, modulating them and their regulation – which is driven by multiple regulons and operons – both qualitatively and quantitatively.

Second, we tacitly assumed that genetic change causes the metabolic differences leading to cross-feeding. However, phenotypic plasticity may also be involved, especially for the consumer strain. Once a producer strain has evolved and excreted a new metabolite, other individuals in the population may sense the new metabolite and respond accordingly; perhaps through a change in gene regulation that does not require mutations. Such plasticity may be important for yet-to-be-discovered instances of cross-feeding, but we know that it is not solely responsible for experimentally characterized cross-feeding interactions. For example, mutations in the regulatory region of gene acs, which expresses the enzyme acetyl CoA synthetase needed for acetate uptake, occurred in all acetate consumer strains that evolved in parallel chemostat experiments (Treves et al., 1998). Such parallel evolution indicates that the mutations may be required for the evolution or maintenance of cross-feeding interaction (Blount et al., 2018). Similarly, three mutations are required when the acetate consumer strain that evolved in batch experiments is to invade and coexist with the acetate producer strain (Plucain et al., 2014).

There may be several reasons why only acetate and glycerol cross-feeding have been observed experimentally, even though multiple other carbon sources may be just as likely to evolve cross-feeding. First, we cannot exclude that other cross-feeding interactions did evolve but went undetected, because detecting cross-feeding requires extensive genetic and biochemical analysis. In addition, cross-feeding strains must reach a sufficiently high population frequency to become detectable. In previous work (San Roman and Wagner, 2018) we showed that about half of the metabolites we study here would

support lower community biomass than acetate cross-feeding. Thus, the frequencies of some cross-feeding strains in a population may be low and hard to detect. Second, cross-feeding interactions may be transient. Given sufficient time, a generalist strain that combines maximum glucose uptake with the ability to recover a secondary metabolite could evolve (Rosenzweig et al., 1994; Le Gac et al., 2012). Evolution experiments in environments alternating between pairs of carbon sources showed that such generalists evolve when both carbon sources are metabolized in similar ways, whereas specialists evolved when carbon sources are metabolized differently (Sandberg et al., 2017). Based on this observation, one would expect that a generalist might replace a cross-feeding polymorphism if glucose and the cross-fed metabolite are metabolically similar. Third, the evolution of acetate and glycerol cross-feeding in the chemostat experiments may have been facilitated by the initial genotype (Kinnersley et al., 2014; Treves et al., 1998). The reason is that the ancestral genotype in these experiments harbored regulatory mutations that prevented cells from recovering excreted acetate and overexpressed the glycerol regulon.

In this contribution, we only studied 58 cross-feeding interactions that can evolve in a single minimal glucose-limited environment. Hundreds or thousands of other cross-feeding interactions can evolve in minimal environments with different primary carbon sources (San Roman and Wagner, 2018), and many more interactions are conceivable in complex environments. To validate such cross-feeding predictions through long term evolution experiments that directly assay for such cross-feeding remains an important task for future work. However, even if only a small fraction of these interactions can be experimentally verified, cross-feeding will emerge as an important source of biodiversity in unstructured and homogeneous environments.

3.5 Methods

3.5.1 Flux balance analysis (FBA)

Flux balance analysis (FBA) is a computational method to predict metabolic fluxes – the rate at which chemical reactions convert substrates into products – of all reactions in a genome-scale metabolic network (Orth et al., 2010). It has been successfully used in many applications, for example to study bacterial growth in different environments (Ibarra et al., 2002) or in response to gene deletions (Edwards and Palsson, 2000).

FBA requires information about the stoichiometry of chemical reactions in a metabolic network. It makes two central assumptions. The first is that cells are in a metabolic steady-state. The second is that cells effectively optimize some metabolic property such as biomass production (growth). Additional constraints can be incorporated into the optimization problem that FBA solves, in order to account for the thermodynamic and enzymatic properties of a network's biochemical reaction. The

optimization problem that FBA solves can be formalized as a linear programming problem (Orth et al., 2010; Varma and Palsson, 1994b) in the following way:

$$\begin{aligned} &Maximize\ v_{growth} \\ &s.t.\ Sv = 0 \\ &l_i \leq v_i \leq u_i \end{aligned} \tag{3.1}$$

Here, S is the stoichiometric matrix, a matrix of size $m \times r$ that mathematically describes the stoichiometry of the modeled network's metabolic reactions. The integer m denotes the number of metabolites, and r denotes the number of biochemical reactions in the network. These reactions include all known metabolic reactions that take place in an organism, which are called internal reactions. They also include reactions that represent the exchange (import or export) of metabolites with the external environment. Furthermore, they include a biomass growth reaction, which is a "virtual" reaction that reflects in which proportion biomass precursors are incorporated into the biomass of the modeled organism (Ibarra et al., 2002; Varma and Palsson, 1994b; Orth et al., 2011). Each entry S_{ij} of the stoichiometric matrix contains the stoichiometric coefficient with which metabolite i participates in reaction j. The vector v is a vector (of size r) whose entries v_i represent the metabolic flux through reaction i in the network. v_{growth} specifies the flux through the biomass growth reaction. Fluxes through biochemical reactions are restricted by lower and upper bounds that constrain the flux through each reaction in the network. These bounds are given by the variables l and u, respectively, which are vectors of size r.

3.5.2 Identification of the flux distribution that characterizes the ancestral strain

Cross-feeding interactions evolve when *E. coli* cells are grown in a glucose-minimal chemostat environment (Helling et al., 1987). In this experiment, the ancestral strain, i.e., the strain present at the beginning of the experiment, is able to grow at a rate equal to the dilution rate of the chemostat (0.2 h^{-1}) while consuming the only carbon source present (glucose).

Mirroring these conditions, we used the genome scale metabolic model of *E. coli* iJO1366 (Orth et al., 2011) and simulated a minimal chemical environment containing glucose as the sole source of carbon. We set glucose consumption to a maximum of 10 $mmol\,gDW^{-1}\,h^{-1}$, an arbitrary value based on typical glucose uptake rates in *E. coli* (Varma and Palsson, 1994b; Orth et al., 2011). We assumed that ammonium, calcium, chloride, cobalt, copper, iron, magnesium, manganese, molybdate, nickel, oxygen, phosphate, potassium, protons, sodium, sulphate and zinc are present in non-limiting amounts. Moreover, we assumed that our (simulated) ancestral strain was able to grow at a dilution rate of 0.2 h^{-1}. Then, we used parsimonious Flux Balance Analysis (pFBA) (Lewis et al., 2010) to identify the flux distribution that best describes the ancestral strain.

pFBA is a variation of FBA . It embodies the hypothesis that organisms have evolved the ability to grow at a maximally possible rate but at a minimal cost, for example, in the form of enzyme expression. Its predictions are more accurate than those obtained with traditional FBA (Lewis et al., 2010). pFBA identifies the flux distribution that satisfies a given growth rate (for example growth at the chemostat dilution rate) while minimizing the sum of all fluxes - a proxy for the total energetic cost of expressing enzymes and transporters. This optimization can be formalized as:

$$\begin{aligned} &Minimize \sum |a_i| \\ &s.t.\ Sa = 0 \\ &l_i \leq a_i \leq u_i \\ &a_{growth} = 0.2 \end{aligned} \tag{3.2}$$

We performed pFBA in python, using the cobrapy package (Ebrahim et al., 2013).

Using pFBA we identified a flux distribution (a for ancestral) that satisfies the growth rate and glucose consumption constraints while minimizing the total flux. With this flux distribution, the *E. coli* metabolism supports growth at the dilution rate of 0.2 h^{-1}, and completely oxidizes 2.1 $mmol\, gDW^{-1}\, h^{-1}$ of glucose consumed, excreting carbon dioxide as the sole carbon containing metabolite.

3.6 Supplementary material

3.6.1 Comparing computational predictions and experimental data

Thorough experimental analysis performed on the ancestral and acetate/glycerol cross-feeding *E. coli* strains provided extensive information on gene expression differences among the strains (Kinnersley et al., 2009). Transcriptional profiling of the evolved strains (acetate and glycerol cross-feeding strains) in monoculture revealed 181 genes whose expression was significantly altered in the evolved strains relative to the ancestral strain (Kinnersley et al., 2009). Approximately half (93) of these genes were classified as metabolic genes (using MultiFun (Serres and Riley, 2000)). Forty-nine genes were up- or down-regulated relative to the ancestral strain, and showed similar expression changes in the acetate and glycerol cross-feeding strains. Forty-four genes showed different expression changes among the cross-feeding strains.

Although our analysis is based on metabolic fluxes and not gene expression, and although the two quantities may show very limited association, for example because of post-translational regulation and differences in enzyme half-lives (Futcher et al., 1999; Greenbaum et al., 2003a; Ideker et al., 2001; Washburn et al., 2003; Vogel and Marcotte, 2012; Guimaraes et al., 2014; Gygi et al., 1999), we

wished to analyze to what extent computationally predicted metabolic changes agree with published experimental gene expression data.

To compare flux predictions with gene expression data, we used the Gene-Protein-Reaction association (GPR) map available for the metabolic model of *E. coli* iJO1366 (Orth et al., 2011). The relationship between genes and enzymatic reactions is only one-to-one in the simplest case, where one gene product catalyzes one reaction. Alternatively, a gene product may catalyze more than one reaction; the products of multiple genes may be needed to catalyze one reaction; or the products of different genes may catalyze the same reaction.

Gene expression data is available for 72 genes included in the genome scale metabolic reconstruction we used for this analysis (iJO1366) (Kinnersley et al., 2009, 2014). The products of most of these genes are associated with few reactions. Exceptions include ompF (Blattner id b0929), tesA (b0494), and ptsH (b2415), which are associated with 271, 21, and 18 reactions. Gene ompF expresses an outer membrane porin which is involved in many transport reactions; tesA expresses acyl-CoA thioesterase I, an enzyme involved in the biosynthesis of unsaturated fatty acids; and ptsH expresses the phosphocarrier protein HPr involved in the phosphotranspherase system (PTS). In total, these 72 genes are associated with 457 reactions.

Four strains evolved and coexisted for many generations in the glucose minimal chemostat experiment (Helling et al., 1987): an acetate consumer (strain CV101), an acetate and glycerol producer (strain CV103), and two glycerol consumers (strains CV115 and CV116). With the aim to disentangle the metabolic changes required for acetate and glycerol production and consumption, we computationally predicted four different flux distributions, i.e., a distribution for an acetate producer, an acetate consumer, a glycerol producer, and a glycerol consumer. We then compared the experimentally observed gene expression changes with the computationally identified flux changes relative to the ancestor. Specifically, we compared (i) the computationally predicted flux changes for the acetate producer strain with the gene expression changes observed experimentally for strain CV103 (an acetate and glycerol producer); (ii) the computationally predicted flux changes for the acetate consumer strain with the gene expression changes observed experimentally for strain CV101 (an acetate consumer strain); (iii) the computationally predicted flux changes for the glycerol producer strain with the gene expression changes observed experimentally for strain CV101 (which produces both acetate and glycerol); and (iv) the computationally predicted flux changes for the glycerol consumer strain with the gene expression changes observed experimentally for strains CV115 and CV116 (two glycerol consumer strains).

For each pair of computationally predicted and experimentally observed changes we performed the following analyses. First, we classified all 72 metabolic genes in the metabolic network of *E. coli* as up-regulated, down-regulated or unchanged relative to the ancestor, using available gene expression

data from (Kinnersley et al., 2009, 2014) for the strains listed above. Then, we classified all (1295) remaining genes included in the model of *E. coli* which were not part of these 72 genes as genes with unchanged expression. Supplementary 3.5 A to E Figs shows in the form of grey bars the number of genes that were classified as up-regulated, down-regulated or unchanged for every experimentally observed strain.

Second, for every gene classified as up-regulated, we examined if at least one of the associated reactions showed a flux increase in the evolved strain (flux e_i) relative to the ancestral strain (flux a_i, $|e_i| > |a_i| + 0.001$). If so, we considered that the experimentally observed gene expression change agreed with the computationally predicted flux change. We applied the same procedure to genes classified as down-regulated, requiring that at least one of the associated reactions showed a flux decrease ($|e_i| < |a_i| - 0.001$) to consider that experimental data agreed with the computational prediction. For genes classified as unchanged, we only assumed agreement between experiment and computational prediction when all reactions associated with a gene showed the same flux in the evolved and the ancestral strain ($|a_i| - 0.001 < |e_i| < |a_i| + 0.001$).

We repeated the analysis just described for computational predictions based on RooM, MoMA, minimization of changed reaction subsets and RooM-het. The results are shown in supplementary 3.5 Fig. Green bars in supplementary 3.5 A to E Figs show the number of genes showing agreement between the experiment and computation (true positives). Orange bars show the number of false positives, i.e., genes computationally predicted to be up-regulated, down-regulated, or unchanged, but where this prediction was not supported by experimental data.

The same data is summarized in panel F of supplementary 3.5 Fig, a table whose entries show the number of true positives and false positives, respectively, separated by a slash ('/'). We performed a Fisher exact test to test the null hypothesis that the number of genes correctly predicted to be up-regulated, down-regulated or unchanged cannot be attributed to chance alone. We found predictions to be significant (p-value<0.05, denoted with *) and very significant (p-value<0.01, denoted with **) for all consumer strains, irrespective of the computational method used to identify the strain.

The data shows that the agreement between experimental data and computational predictions is very limited, which is unsurprising, given that metabolic flux is affected by multiple factors besides mRNA expression level. In addition, computational predictions are even less accurate for producer strains than for consumer strains. One possible reason is that the strain found experimentally to excrete acetate (CV103) also excretes glycerol, while we computationally modeled two separate producer strains excreting either acetate or glycerol.

3.6.2 Alternative methods used to identify the cross-feeding strains.

In this supplementary file we introduce alternative methods that we used to identify the flux distribution of the cross-feeding strains. We also present a comparison of the results obtained with these methods and the results presented in the main text obtained with RooM.

Minimization of Metabolic Adjustments (MoMA)

MoMA, like RooM, was first described (Segrè et al., 2002) to predict metabolic changes after genetic perturbations such as gene deletion. It is based on the assumption that after a genetic perturbation, the flux change in a metabolic network relative to a wild-type will be minimal. It identifies a flux distribution that satisfies a set of constraints while minimizing the distance to a flux distribution, just as RooM does. The difference between RooM and MoMA consists in the distance metric used: While in RooM the number of reactions with a significant flux change relative to a reference flux distribution is minimized, in MoMA the Euclidean distance to the reference flux distribution is minimized. The quadratic programming problem solved in MoMA can be written as:

$$\begin{aligned} &Min\ \sqrt{\sum (a_i - e_i)^2} \\ &s.t.\ Se = 0 \\ &l_i \leq e_i \leq u_i \end{aligned} \tag{3.3}$$

Here, S is the m×r stoichiometry matrix and l_i and u_i are lower and upper bounds, respectively, that constrain the flux through each reaction i in the network. These bounds reflect thermodynamic and capacity constraints on a reaction. The flux distribution a corresponds to the reference flux distribution, which for our analysis equals the flux distribution of the ancestral strain obtained with pFBA, as explained in the main text.

We repeated the analysis presented in the main text, identifying every producer and consumer strain with MoMA instead of RooM. We found that the mean ancestor-producer distance was 10.3±7.7 $mmol\, gDW^{-1}\, h^{-1}$ (supplementary 3.6A Fig). The distances between the ancestor and the acetate and glycerol producer strains were 3.3 $mmol\, gDW^{-1}\, h^{-1}$ and 2.9 $mmol\, gDW^{-1}\, h^{-1}$, respectively. When ranking producer strains according to increasing distance to the ancestor, the acetate producer had rank eleven, while the glycerol had rank six. The ancestor-consumer mean distance was 27.8±41.6 $mmol\, gDW^{-1}\, h^{-1}$ (figure S3_fig (B)). The acetate consumer strain occupied rank 47, with a distance of 26.5 $mmol\, gDW^{-1}\, h^{-1}$. The glycerol consumer strain ranked third, with a distance of 8.6 $mmol\, gDW^{-1}\, h^{-1}$.

We used the sum of the ancestor-producer and ancestor-consumer distances as a proxy for the likelihood of cross-feeding interactions to evolve (supplementary 3.6C Fig). We observed that 34

cross-feeding interactions can evolve more easily than the one involving acetate. In contrast, only two cross-feeding interactions can evolve more easily than glycerol cross-feeding, those involving the metabolites dihydroxyacetone (dha) and D-gluconate (glcn). Even though MoMA and RooM perform different optimizations to identify the flux distributions of the cross-feeding strains and even though the distances obtained differ in absolute values, we observed a high statistical association (Spearman's r= 0.86, P=6.1e-18, n=58) between the predictions of both methods with respect to metabolite rank, i.e., the likelihood with which metabolite cross-feeding arises.

Minimizing the number of reaction subsets with a significant change in flux

Regulatory mutations may be important for causing the flux changes that allow cross-feeding, and any one regulatory mutation may affect the flux through multiple reactions simultaneously. Bearing this in mind, we wanted to quantify the number of flux changes not in terms of individual reactions (as in RooM) but in terms of subsets of reactions which are co-regulated and whose flux may be simultaneously affected by a single regulatory change. To this end, we first identified all subsets of fully coupled reactions. (A set of reactions is fully coupled if the flux through one reaction constrains the flux through all reactions in the set to a scalar multiple of the specified flux (Larhlimi et al., 2012).These fully coupled reactions tend to be co-regulated (Notebaart et al., 2008). We used the tool F2C2 (Larhlimi et al., 2012) to identify such reactions. We included all reactions that were fully coupled in one subset of reactions. With this approach, we partitioned the 2583 reactions from the *E. coli* metabolic model iJO1366 into 1118 subsets of fully coupled reactions.

We then performed an optimization similar to RooM, in which a flux distribution that fulfills a set of constraints is identified, while the number of subsets of reactions that change their flux with respect to the ancestral flux distribution is minimized. The optimization problem can be written as:

$$\begin{aligned} & Min \sum f_i^{subset} \\ & s.t.\ Se = 0 \\ & f_i^{subset} \in 0, 1 \\ & l_i \leq e_i \leq u_i \\ & e_i - f_i^{subset}(u_i - (a_i + \beta)) \leq a_i + \beta \\ & e_i - f_i^{subset}(l_i - (a_i - \beta)) \geq a_i - \beta \end{aligned} \tag{3.4}$$

Once again, here S is the $m \times r$ stoichiometry matrix and l_i and u_i are lower and upper bounds respectively, that constrain the flux through each reaction in the network according to thermodynamic and capacity constraints. f_i^{subset} is a binary variable. It takes a value of 1 if reaction subset i shows a substantial change in flux e_i relative to the reference flux a_i and zero otherwise. β specifies the amount of flux change that is considered substantial, and as in RooM, we used a value of β=0.001.

We used this method to identify all potential cross-feeding strains. We found that producers differed from the ancestor on average in 50±10 reaction subsets (supplementary 3.6D Fig) while consumers differed from the ancestor on average in 55±13 reaction subsets (supplementary 3.6E Fig). We also used the sum of the ancestor-producer and ancestor-consumer distances obtained with this method to estimate the likelihood of the various cross-feeding interactions to evolve. For acetate and glycerol cross-feeding 92 and 83 subsets of reactions require a flux change, respectively. Cross-feeding of 17 metabolites is easier to evolve than acetate cross-feeding, whereas cross-feeding of only 6 metabolites is easier to evolve than glycerol cross-feeding.

Although the results presented in the main text considered all reactions, whereas the method used here considered only reactions subset, the methods produced highly concordant predictions about metabolite rank,i.e., about the likelihood that metabolite cross-feeding arises (Spearman's r= 0.98, P=2.4e-44, n=58).

3.6.3 Model to quantify the probability of cross-feeding evolution.

The evolution of the consumer strain may not be independent of that of the producer. For example, if the producer is not present in the population by the time the consumer evolves, the consumer will go extinct because it lacks a carbon source on which to thrive. Therefore, the sum of ancestor-producer and ancestor-consumer distances may not be the most accurate proxy for the likelihood that cross-feeding evolves. In this section we develop a more adequate model that does not consider producer and consumer origins as independent. We also show that the predictions of this more complex model are virtually identical to those of the simpler model of the main text.

One fundamental assumption of our work is that metabolic ancestor-producer and ancestor-consumer distances are proportional to the amount of metabolic or genetic change that is necessary to give rise to a producer or a consumer strain for any one carbon source. In addition, we assume that these distances are inversely proportional to the likelihood that a producer or consumer strain emerges. In a reflection of these assumptions, we transformed metabolic distances (d) into probabilities that producers (consumers) emerge in the population after one generation (P). To this end, we took a phenomenological approach and assumed that the probability that an evolved (producer or consumer) strain originates in the population from the ancestor would have a maximal value of P=1 if the metabolic distance between ancestor and the evolved strain had a minimal value of zero. This probability would monotonically approach the smallest value of zero as the metabolic distance became larger. We used the following simple function to transform metabolic distances into probabilities of evolution:

$$P = \frac{1}{1+d} \tag{3.5}$$

Equation 3.5 predicts, for example, that the producer of dihydroxyacetone (dha), which requires the

fewest flux changes (19 reactions, figure 3.2A), is the one with the highest probability to emerge in the population (P=0.050). In addition, the dihydroxyacetone (dha) consumer strain is also the consumer strain with the highest probability to emerge (P=0.11), because it requires only 8 reactions (figure 3.2A) to change their flux relative to the ancestral strain. Conversely, the strain producing L-alanine-D-glutamate-meso-2,6-diaminoheptanedioate-D-alanine (LalaDgluMdapDala) has the lowest probability of emerging among all producer strains (90 flux changes, P=0.011), and the strain consuming formate (for) has the lowest probability among all consumer strains (99 reaction changes, P= 0.010).

We used these probabilities to model the evolution of cross-feeding. For simplicity, we assumed that generations are non-overlapping and that at every generation four scenarios are possible: the producer and consumer strains arise simultaneously, which occurs with probability $P_{Producer \wedge Consumer}$; only the producer arises (probability $P_{Producer \wedge (\neg Consumer)}$; only the consumer arises (probability $P_{(\neg Producer) \wedge Consumer}$; or neither producer nor consumer arises (probability $P_{(\neg Producer) \wedge (\neg Consumer)}$. For cross-feeding to evolve after the first generation, producer and consumer need to arise simultaneously. Therefore the probability of observing cross-feeding after the first generation is given by $P(cf, t_1) = P_{Producer \wedge Consumer}$. From the second generation onwards cross-feeding may arise by simultaneous evolution of producer and consumer (as in the first generation) or by evolution of a consumer after the evolution of a producer. The probability of evolution of cross-feeding in generation t (with $t > t_1$) can be written as:

$$
\begin{aligned}
&P_{cf,t} = P_{(\neg Producer)^{t-1}} P_{Producer \wedge Consumer} + \\
&\sum_{m=2}^{t} \sum_{j=0}^{t-m} [(P_{(\neg Producer)})^{m-2} (P_{Producer \wedge (\neg Consumer)})^{j+1} (P_{(\neg Producer) \wedge (\neg Consumer)})^{t-m-j} \\
&\frac{(t-m)!}{j!(t-m-j)!} P_{Consumer}]
\end{aligned}
\tag{3.6}
$$

Where $P_{(\neg Producer)}$ (with $P_{(\neg Producer)} = P_{(\neg Producer) \wedge Consumer} + P_{(\neg Producer) \wedge (\neg Consumer)}$) is the probability that the producer does not emerge in a given generation. $P_{Consumer}$ (with $P_{Consumer} = P_{Producer \wedge Consumer} + P_{(\neg Producer) \wedge Consumer}$) is the probability of the consumer to emerge in a given generation.

Defining p and c as the probabilities that producer and consumer evolve in the population, then for a population composed of N cells (N=2x1010 in (Helling et al., 1987)) we can write the following probabilities:

$$
\begin{aligned}
&P_{\neg Producer} = (1-p)^N \\
&P_{Producer} = 1 - P_{(\neg Producer)} = 1 - (1-p)^N \\
&P_{(\neg Consumer)} = (1-c)^N \\
&P_{Consumer} = 1 - P_{(\neg Consumer)} = 1 - (1-c)^N \\
&P_{(\neg Producer)\wedge(\neg Consumer)} = (1-(p+c))^N \\
&P_{Producer\wedge Consumer} = \sum_{r_P=1}^{N-1} \sum_{r_C=1}^{N-r_P} \frac{N! p^{r_P} c^{r_C} (1-p-c)^{N-r_P-r_C}}{r_P! r_C! (N-r_P-r_C)!} \\
&1 + (1-p-c)^N - (p+(1-p-c))^N - (c+(1-p-c))^N \\
&P_{Producer\wedge(\neg Consumer)} = P_{Producer} - P_{Producer\wedge Consumer} \\
&P_{(\neg Producer)\wedge Consumer} = P_{Consumer} - P_{Producer\wedge Consumer}
\end{aligned}
\tag{3.7}
$$

We used equation 3.5 to estimate the probabilities of the producer $P_{Producer}$ and consumer $P_{Consumer}$ strains to evolve in a population. With these probabilities, we were able to calculate the probabilities of producer and consumer to evolve per cell in the population (p and c). With p and c in hand, we determined the remaining probabilities needed to calculate the probability of cross-feeding: $P_{(\neg Producer)\wedge(\neg Consumer)}$, $P_{Producer\wedge Consumer}$, $P_{Producer\wedge(\neg Consumer)}$ and $P_{(\neg Producer)\wedge Consumer}$.

Supplementary 3.7A Fig plots the predicted cumulative probability ($\sum_t P(cf,t)$) that cross-feeding emerges against time (in generations) for the evolution of each of the 58 possible cross-feeding interactions. Orange and green lines show the predictions for acetate and glycerol respectively. Supplementary 3.7B Fig compares this prediction to that of the simpler model of the main text, where the sum of the ancestor-producer and ancestor-consumer distances are used as proxies for these probabilities. The two predictions are highly concordant (Spearman's r=0.99, P=9.7e-74, n=58).

3.6.4 Supplementary figures

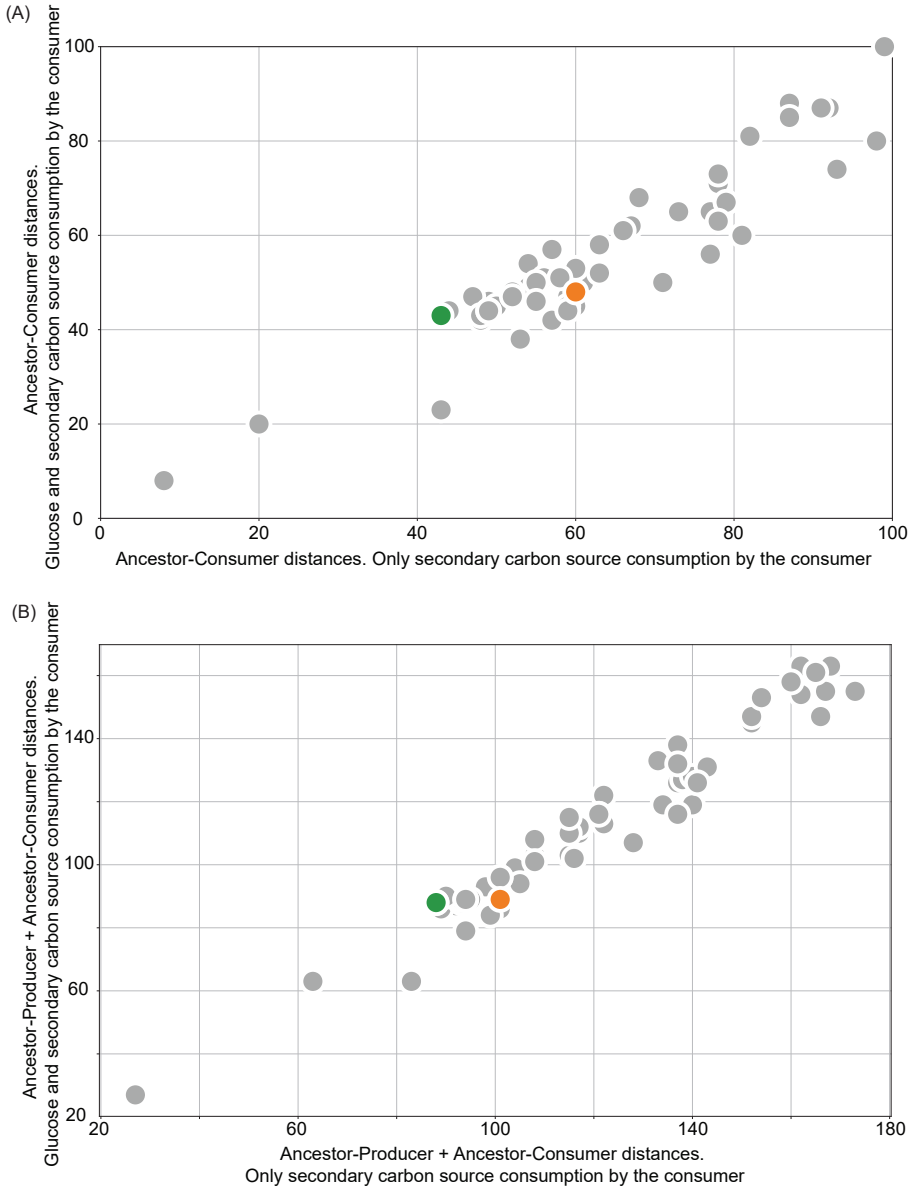
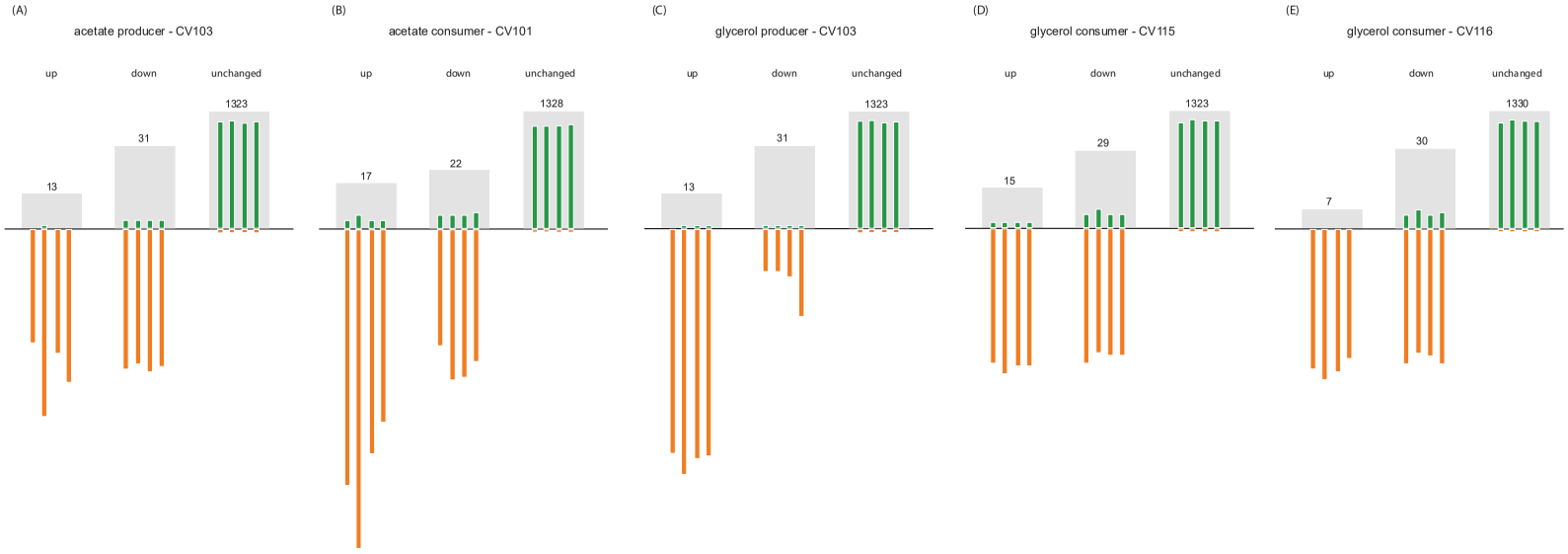


Figure 3.4: Comparison of ancestor-consumer distances and B) the sum of ancestor-producer and ancestor-consumer distances when the consumer strains consume (A) only the specific secondary carbon source or (B) both glucose and the secondary carbon source The x-axes show the ancestor-consumer and total distances obtained when the consumer strains cannot consume glucose but only the specific secondary carbon source (see also figures 3.2 A and B). The y-axes show the same distances but obtained when the consumer strain consumes 1 $mmol\, gDW^{-1}\, h^{-1}$ of glucose and the respective secondary carbon source in amounts that allow growth at 0.2 h-1. The ancestor-producer distances used to calculate the total distances shown in (B) are those shown on the x-axis of figure 3.2A. Every grey circle represents one of 56 metabolites that can be cross-fed. Acetate and glycerol are shown as orange and green circles, respectively. Even though the ancestor-consumer distances change when the consumer strains consume glucose in addition to their specific secondary carbon source, the main conclusion of this work do not change: Multiple changes are required for any cross-feeding interaction to evolve, and cross-feeding of multiple metabolites may evolve with higher likelihood than acetate and glycerol cross-feeding.



(F)

	up-regulated				down-regulated				unchanged			
	RooM	MoMA	subsets	RooM-het	RooM	MoMA	subsets	RooM-het	ROOM	MoMA	subsets	ROOM-het
acetate producer - CV103	0/43	1/71	0/47	0/58	3/53	3/51	3/54	3/52	1212/39	1223/37	1196/36	1213/38
acetate consumer - CV101	3/97	* 5/121	3/85	3/73	** 5/44	** 5/57	** 5/56	** 6/50	** 1164/27	** 1165/25	1167/38	** 1182/28
glycerol producer - CV103	0/85	1/93	1/87	1/86	1/16	1/16	1/18	1/33	1219/39	1226/37	1201/37	* 1210/36
glycerol consumer - CV115	2/51	2/55	2/52	2/47	** 5/51	** 7/47	** 5/48	** 7/50	** 1195/32	** 1230/33	** 1215/34	** 1212/32
glycerol consumer - CV116	0/53	0/57	0/54	9/49	** 5/51	** 7/47	** 5/48	** 6/51	** 1200/27	** 1235/28	** 1221/28	** 1217/27

Figure 3.5 ***(previous page)*****: Comparison of gene expression data of the cross-feeding strains experimentally evolved in glucose minimal chemostats with computationally predicted flux changes.** The light grey areas in figures (A) to (E) show the number of genes found to be up-regulated, down-regulated, or unchanged in expression for the experimentally observed acetate producer CV103, the acetate consumer CV101, the glycerol producer CV103, and the two glycerol consumer strains CV115 and CV116. The numbers above the grey bars add up to the total number of genes included in the metabolic model of *E. coli* iJO1366 (1367). Green and orange bars show the genes predicted to be up-regulated, down-regulated or unchanged, based on flux changes for reactions associated with these genes, as predicted by (from left to right) RooM, MoMA, minimizing reaction subsets, and RooM-het. Green bars (overlapping the grey area) indicate the number of genes correctly predicted to be up-regulated, down-regulated, or unchanged (true positives). Orange bars indicate the number of genes computationally predicted but not experimentally observed to be up-regulated, down-regulated, and unchanged (false positives). (F) Summarizes the data shown in (A) to (E). For each strain (rows) and each gene category and prediction method (columns), the two numbers separated by a dash indicate the number of true positives and false positives. '*' indicates $p<0.05$, and '**'$p<0.01$, based on a Fisher's exact test of the null-hypothesis that the number of genes correctly predicted to be up-regulated, down-regulated, or unchanged can be attributed to chance alone.

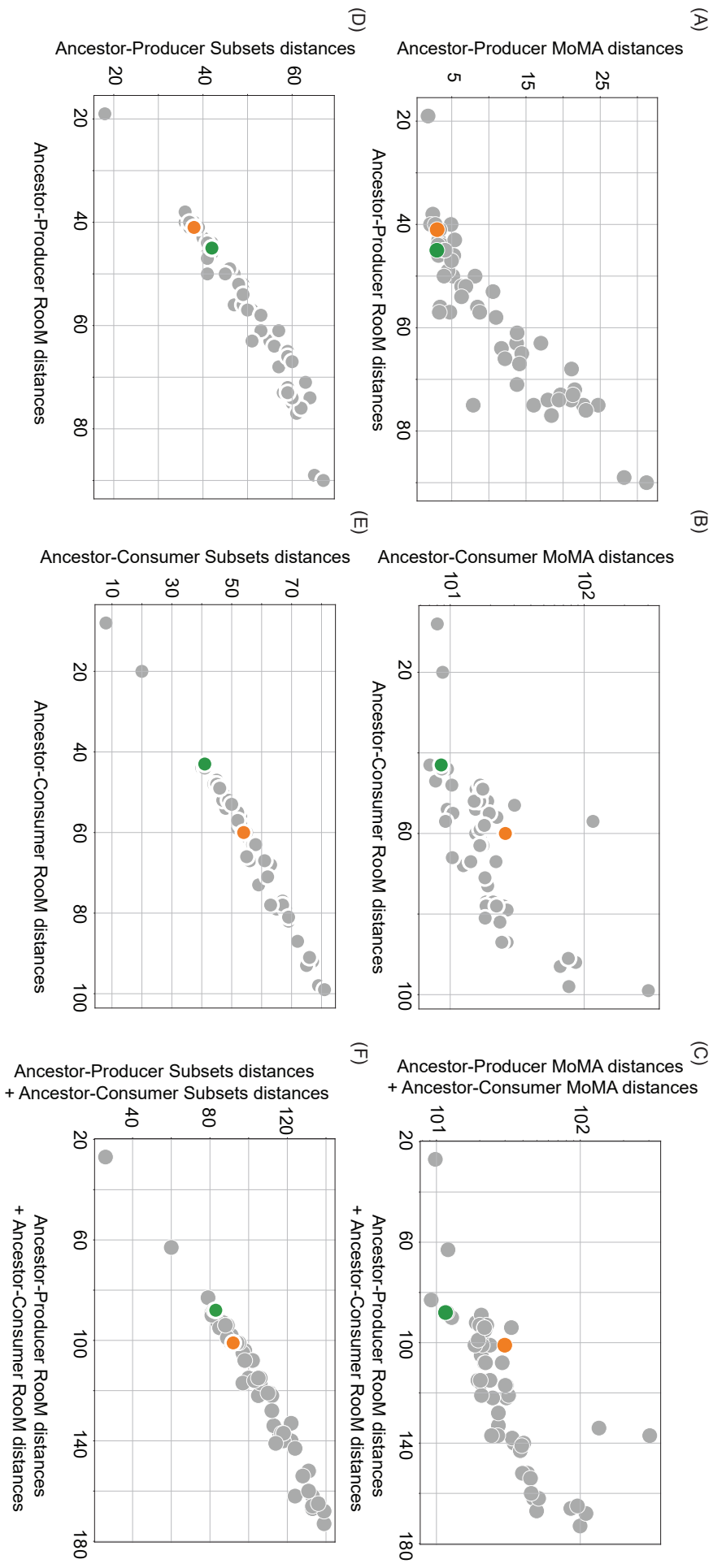

(A)
Ancestor-Producer MoMA distances
Ancestor-Producer RooM distances
(B)
Ancestor-Consumer MoMA distances
Ancestor-Consumer RooM distances
(C)
Ancestor-Producer MoMA distances
+ Ancestor-Consumer MoMA distances
Ancestor-Producer RooM distances
+ Ancestor-Consumer RooM distances
(D)
Ancestor-Producer Subsets distances
Ancestor-Producer RooM distances
(E)
Ancestor-Consumer Subsets distances
Ancestor-Consumer RooM distances
(F)
Ancestor-Producer Subsets distances
+ Ancestor-Consumer Subsets distances
Ancestor-Producer RooM distances
+ Ancestor-Consumer RooM distances

Figure 3.6 ***(previous page)*****: Comparison of the ancestor-producer, ancestor-consumer and total distances obtained when using different methods to identify flux distributions of cross-feeding strains.** In figures (A) to (C) flux distribution distances predicted by MoMA (on the y-axis) are compared with distances predicted by RooM (on the x-axis). The ancestor-producer distance, the ancestor-consumer distance, and the total distance (sum of ancestor-producer and ancestor-consumer distances) are shown in panels (A) to (C), respectively. Every grey circle represents one of 56 metabolites that can be cross-fed. Acetate and glycerol are shown as orange and green circles. Panels (D) to (F) are analogous to (A) to (C), but their y-axes show the distances predicted when minimizing the number of co-regulated reaction subsets that change expression.

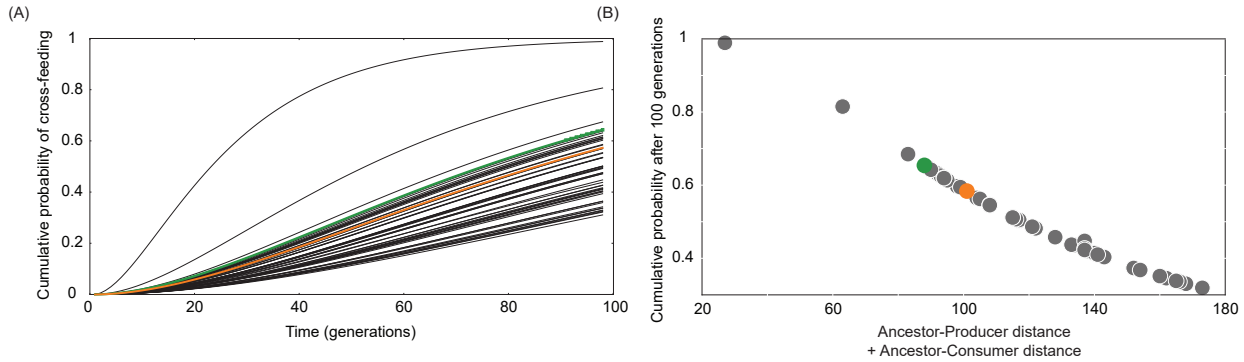


Figure 3.7: Estimates of likelihood of cross-feeding to evolve considering the evolution of producer and consumer strains are not independent events. (A) The cumulative probability for the evolution of cross-feeding interactions as calculated with equation 3.6 from supplementary 3.6.3 Text is plotted against time (in generations). Every grey line corresponds to a prediction for a different metabolite subject to cross-feeding. Predictions for acetate and glycerol are shown in orange and green, respectively. (B) Comparison of two proxies of the likelihood that cross-feeding evolves. The x-axis shows the sum of the producer-ancestor and consumer-ancestor distances, which is the proxy used in the main text. The y-axis shows the cumulative probability of cross-feeding to evolve after 100 generations, according to the model from supplementary 3.6.3 Text, which takes into consideration that the evolution of producer and consumer may not be independent events. Every grey circle represents a prediction for a different metabolite subject to cross-feeding. Orange and green circles correspond to predictions for acetate and glycerol, respectively. The two proxies for the likelihood to evolve cross-feeding are highly correlated (Spearman's r=0.99, P=9.7e-74, n=58).

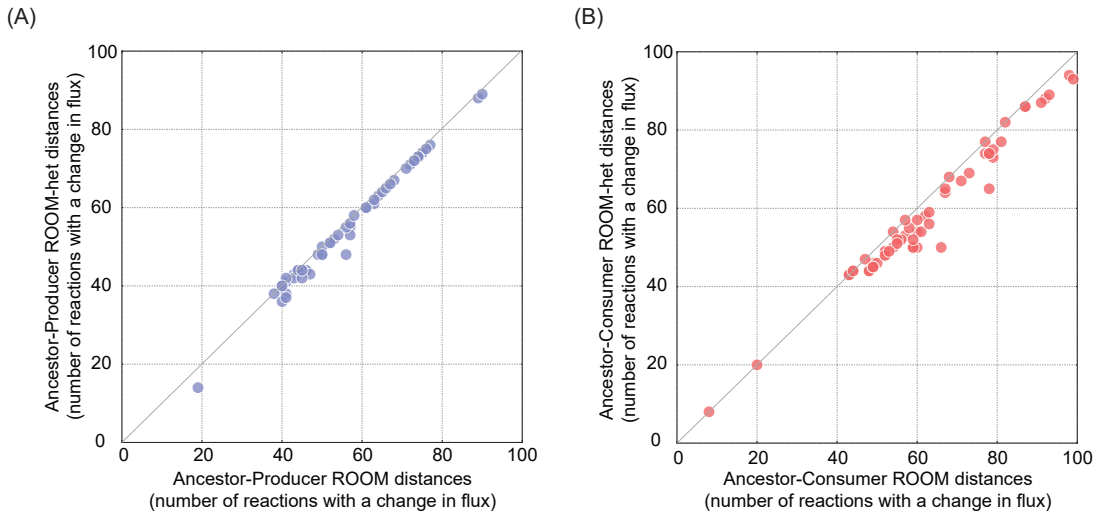


Figure 3.8: RooM-het and RooM distances comparison. (A) Ancestor-producer distances predicted with RooM and RooM-het are shown on the x and y-axes respectively. Every circle corresponds to one produced metabolite. The diagonal line indicates equal distances. (B) As in (A) but for predicted ancestor-consumer distances and for consumed metabolites.

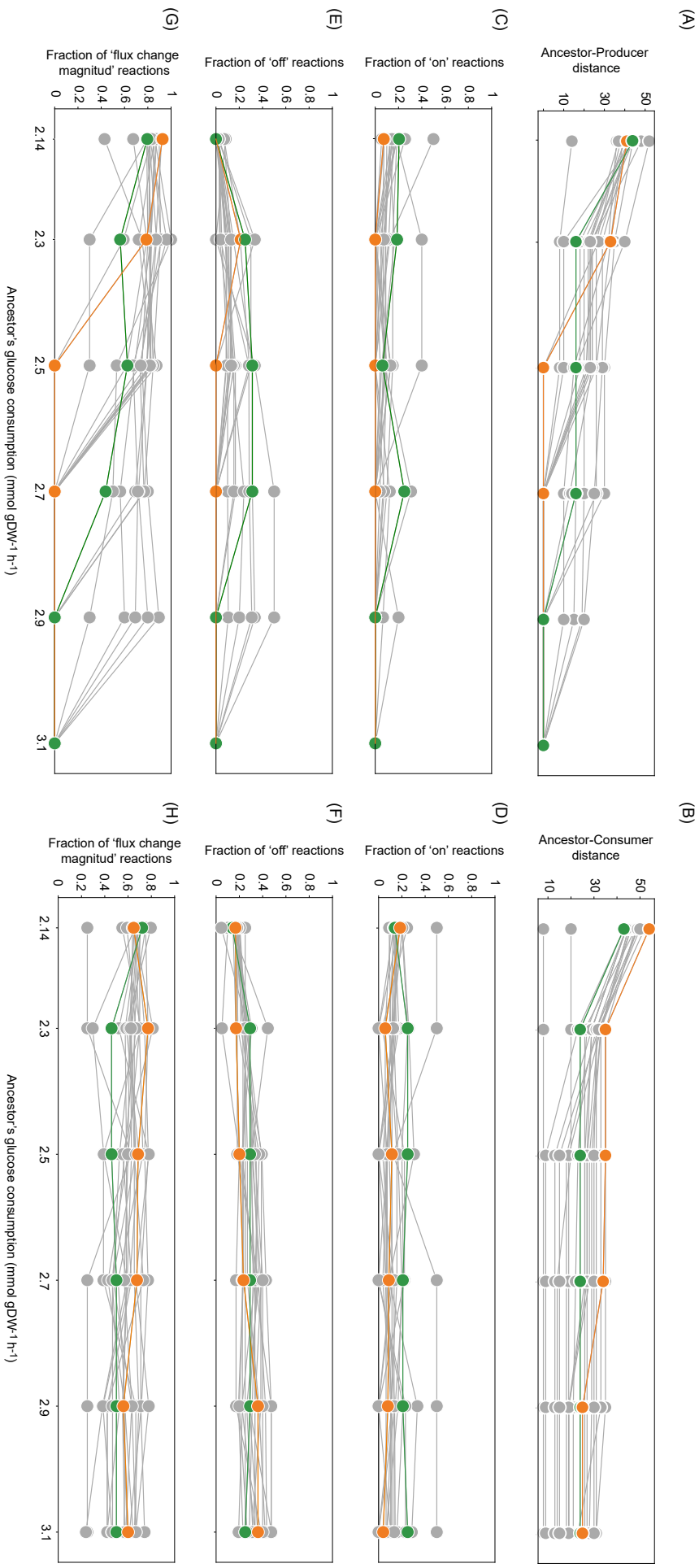

(A)
Ancestor-Producer distance
(B)
Ancestor-Consumer distance
(C)
Fraction of 'on' reactions
(D)
Fraction of 'on' reactions
(E)
Fraction of 'off' reactions
(F)
Fraction of 'off' reactions
(G)
Fraction of 'flux change magnitud' reactions
(H)
Fraction of 'flux change magnitud' reactions
Ancestor's glucose consumption (mmol gDW-1 h-1)
Ancestor's glucose consumption (mmol gDW-1 h-1)

Figure 3.9 ***(previous page)*****:** The x-axes in all panels show the maximal amount of glucose consumed by the ancestor. As a function of this quantity, (A) and (B) show the predicted distance of the ancestor to the producer and consumer, respectively. Among all reactions with a change in flux, (C), (E) and (G) show the fraction of reactions that are turned 'on', 'off' and that change the flux quantitatively in the producer relative to the ancestor, respectively; (D), (F) and (H) are analogous to (C), (E) and (G), but for the reactions that require a flux change in the consumer relative to the ancestor. Each set of six grey circles connected by a grey line corresponds to simulation data for one of the twenty metabolites with a predicted likelihood of being subject to evolve cross-feeding greater than that for acetate when either RooM or RooM-het are performed at the minimal glucose consumption rate of 2.14 $mmol\,gDW^{-1}\,h^{-1}$. Predictions for acetate and glycerol are shown as orange and green circles, respectively.

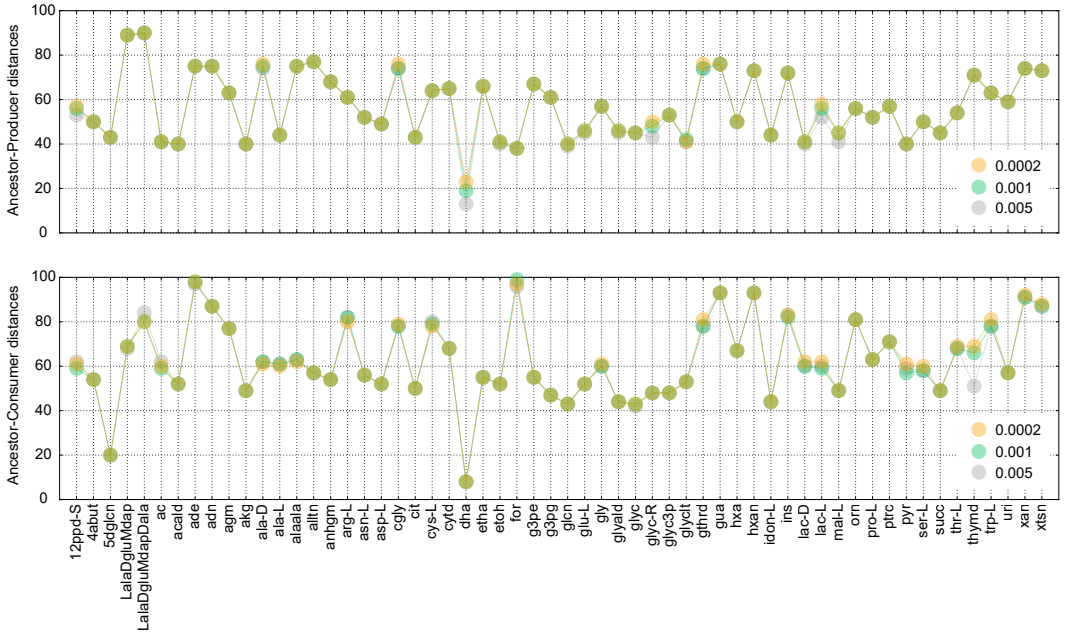


Figure 3.10: The figure shows the ancestor-producer and ancestor-consumer distances (y-axis in the upper and lower panel respectively) predicted by RooM for all 58 cross-feeding strains, using three different values of the parameter beta, which is used in RooM to specify the amount of flux change that is considered substantial (increasing distance in one unit). The data shows that changing beta from its default value (0.001 $mmol\,gDW^{-1}\,h^{-1}$, green circles) to a fifth of this value (0.0002 $mmol\,gDW^{-1}\,h^{-1}$, yellow circles), or to five times this value (0.005 $mmol\,gDW^{-1}\,h^{-1}$, grey circles) has very little effect on the predicted distances.

Chapter 4

From two to many

To be submitted as:

San Roman M. and Wagner A. (2020) Diversity-creating species interactions can shift ecological diversity limits during community assembly.

4.1 Summary

Microbial communities are hugely diverse, but we do not yet understand how species invasions and extinctions drive and limit their diversity. According to the ecological limits hypothesis, available environmental resources limit diversity. According to the diversity-begets-diversity hypothesis, biotic species interactions can create new ecological niches that help increase diversity. To find out whether resource limitations or species interactions may be more important in the assembly of microbial communities, we used metabolic modeling. We represent each microbial species by a metabolic network that harbors thousands of biochemical reactions. Together, these reactions determine which carbon and energy sources a species can consume, and which metabolic by-products – potential nutrients for other species – it can excrete in a given chemical environment. We assemble communities by modeling thousands of individual species invasions in a chemostat-like environment that initially provides only a single source of carbon and energy. We find that early during the assembly process, diversity begets diversity. By-product excretion and cross-feeding interactions produce a rich environment with up to eighty different available carbon sources, which can sustain dozens of species. During later assembly stages, the creation of new niches slows down, existing niches become filled, successful invasions become rare, and species diversity plateaus. Thus, ecological limitations dominate the late assembly process. However, these limitations are not just externally imposed. They result from species interactions that have created new resources during early assembly. A community's diversity ceiling is a joint product of external limitations and internal processes that can raise this ceiling.

4.2 Introduction

Bacterial life on earth is extraordinarily diverse. Our planet is inhabited by an estimated 1.4 to 1.9 million bacterial lineages (Louca et al., 2018), and every gram of soil hosts between 2000 and 18000 distinct such lineages (Daniel, 2005). Understanding how Earth holds all this biodiversity is a fundamental challenge to ecology and evolution. Part of the difficulty in meeting this challenge is that planet earth offers limited resources, which may impose a ceiling on biodiversity. In this vein, a prominent 'ecological limits' model of biodiversity (Rabosky and Hurlbert, 2015; Schluter and Pennell, 2017) posits that rates of diversification should decrease as diversity increases and fills available ecological niches. In contrast, the 'diversity begets diversity' perspective posits that diversity could stimulate further diversification (Whittaker, 1972; Calcagno et al., 2017) as species-species interactions get more complex and novel niches are created (Laland et al., 1999; Erwin, 2008), or as existing niches are partitioned more and more finely as a result of competition (Dieckmann and Doebeli, 1999a; Bailey et al., 2013).

Evidence for either model in bacteria, animals, and plants is mixed. For example, a combination

of field work and phylogenetic analysis of Himalayan songbirds suggests that an observed slowdown in their speciation rate is explained by niche filling. Their species distributions are well explained by resource abundance (Price et al., 2014). Two evolution experiments with *Pseudomonas fluorescens* also support the ecological limits model (Brockhurst et al., 2007; Gómez and Buckling, 2013). In the first experiment, the evolutionary diversification of a strain of *P. fluorescens* (measured as the rate at which distinct morphotypes evolved) was lower when the strain was co-cultured with a community isolated from soil than when it was cultured in isolation (Gómez and Buckling, 2013). In the second experiment, the strain's diversification also slowed down as more strains where included in the co-culture (Brockhurst et al., 2007). Conversely, similar experiments with a different strain of *P. fluorescens* support the 'diversity begets diversity' perspective. These experiments tracked the diversification of a focal lineage of *P. fluorescens* in bacterial communities that harbored from one to eight lineages of *P. fluorescens*. More novel morphotypes evolved when communities were initially more diverse (Jousset et al., 2016). Further support for the 'diversity begets diversity' model comes from experiments with crops and weeds, which show that weed diversity is high whenever crop diversity is high (Palmer and Maurer, 1997). It also comes from data on plant and arthropod diversity on the Canary and Hawaiian islands, where the proportion of endemic species increases with increasing species numbers (Emerson and Kolm, 2005). And it comes from inferences of microbial diversification rates from several taxonomic ratios, such as the species to genus ratio (Madi et al., 2019). In addition, theoretical modeling using adaptive dynamics also suggests that some initial diversity can facilitate further diversification (Calcagno et al., 2017).

The evidence supporting one or other model comes from experiments or field work data where it is difficult to disentangle ecological from evolutionary limitations. We believe this dichotomy of whether diversity fosters or inhibits further diversity should be resolved based on purely ecological processes: in a scenario where we could control the appearance of new species in a community, will the environmental resources constrain community diversity or will species-species interactions promote further diversity? This is what we explore here. We simulated the assembly of microbial communities where each individual species is modeled with a metabolic network which comprises thousands of metabolic reactions needed for an organism's survival in a given chemical environment. The advantage of this approach is that fundamental metabolic traits emerge from first biochemical principles, which are embodied in the metabolic reaction network of an organism. Especially important traits for our purpose include the ability to survive on a given source of carbon and energy, and the ability to excrete specific by-products of metabolism. Traits like these can be computationally predicted with Flux Balance Analysis (FBA, (Orth et al., 2010)), a computational method that can determine the flux of matter through every reaction in a metabolic network when cells are in a metabolic steady-state, and when they grow their biomass at the maximum possible rate given the biochemical reactions they

can catalyze. The predictions of FBA have been experimentally validated (Varma and Palsson, 1994b; Orth et al., 2011).They have also been successfully used to predict growth and by-product secretion of bacteria in different media (Varma and Palsson, 1994b; Edwards et al., 2001; Ibarra et al., 2002).

Metabolic networks have been characterized for hundreds of organisms (Gu et al., 2019). They are highly valuable to understand metabolic biology and evolution, because they reflect an organism's evolutionary history. For the same reason, however, they are of limited use for studies like ours, which aim to understand how community assembly may be affected by specific metabolic properties of the assembled species, such as the number of biochemical reactions any one species harbors, and the number of carbon sources it can utilize. For this purpose, one needs to vary these properties systematically, but in the network of any one organism, they are fixed.

To circumvent this limitation, we started our community assembly not from previously characterized metabolic networks of microbial species. Instead, we randomly sampled metabolic networks from a much larger space of possible metabolisms (Methods), such that each network fulfilled specific requirements. These requirements include the ability to sustain life on specific carbon sources like glucose, while containing an otherwise random complement of chemical reactions drawn from a much larger 'universe' or 'pan-metabolism' of biochemical reactions. We refer to such metabolic networks as random viable networks. We created them with a computational variant of Markov Chain Monte Carlo (MCMC) sampling which allows for a efficient sampling of the large space containing all possible metabolic reaction, that is, the pan-metabolism. For our work, we used a pan-metabolic network comprising 5625 metabolites and 7222 biochemical reactions. Using this approach, we created thousands of random viable networks (see Methods). For the purpose of our analysis, each such network represents a different 'species', and we used these 'species' to simulate the assembly of a community in a well-mixed chemostat-like environment where resources are supplied by the environment at a constant rate.

Our community assembly procedure consists of three iterated steps. First, a random species invades the environment/community (Figure 4.1). This step simulates the evolution or migration of a new species in the community. Second, the environment acts as a filter that selects those species that succeed at persisting in the community. Specifically, we allow a species to persist if it grows faster than all other species in the community on at least one of the available resources (carbon sources). In other words, this second step embodies competition between species in the community. In a third step, the persisting species may change the environment by excreting metabolic by-products. We repeated these three assembly steps for thousands of species invasions in any one community, and simulated hundreds of community assemblies in environments with different amounts of resources and varying strengths of competition.

During community assembly, we found that many persisting species excrete metabolic by-products,

which can sustain species that invade later through cross-feeding – interaction in which one organism consumes the excretion of another. In other words, species create opportunities for cross-feeding, which begets further diversity. Further, we find that communities grow more diverse when competition among species is stronger. Taken together, these two observations show that biotic interactions between species are critical to establishing diverse communities. At the same time, communities show limits in the number of species they accommodate. As community richness increases, the probability that a new species successfully invades a community decreases. The reason is that by-product excretion cannot create new niches ad infinitum. Eventually all new niches have been created, and later invasions mostly fill the existing niches. At this late stage, the assembly dynamics enters an 'ecological limits' regime.

4.3 Results

4.3.1 Modeling community assembly

We used a very simple community assembly strategy inspired by the environmental filter metaphor and trait-based models. In these models life is organized according to traits - properties of a species which are essential for its survival. Typically, species show trade-offs in their traits and a trait that gives a species an advantage in one environment might be disadvantageous in other. Then the environment selects (filters) species according to their traits. These modeling strategy has been used, for example, to predict community composition along environmental gradients (Mcgill et al., 2006; Thakur and Wright, 2017; Allison, 2012; Litchman and Klausmeier, 2008; Laughlin et al., 2012). Here, we model the assembly of microbial communities where the species traits are given by the species ability to grow on a set of carbon sources. In the following paragraphs we describe our modeling procedure in detail.

To begin with, we created a pool of 1000 species, from which we sample individual species for community assembly. Each species is represented by a metabolic network that we require to be viable on a specific carbon source, but that contains an otherwise random complement of biochemical reactions from a much larger universe ('pan-metabolism') of biochemical reactions known to take place in the biosphere. To create this species pool, we used a Markov Chain Monte Carlo (MCMC) sampling procedure, which allows us to create random viable metabolic networks with specific characteristic (see methods). For most of our analysis these characteristics are that, first, all species (metabolic networks)contain the same number of 2583 metabolic reactions as a well-established and widely used metabolic model of *E. coli* (iJO1366, (Orth et al., 2011)). Second, all species are permeable to the same 330 metabolites that *E. coli* cells can import or export (Orth et al., 2011). Third, we require all species to be viable on the same carbon source glucose, a constraint we relax later (supplementary 4.10 Fig). We note that as a result of metabolism's complex and reticulate structure, species required

to be viable on one carbon source are usually also viable on multiple others (Barve and Wagner, 2013).

We used Flux Balance Analysis (see methods) to determine viability for each of our 1000 species on 223 potential carbon sources (see methods), and found that they are viable on an average of 32±10 different carbon sources. We consider the carbon sources on which any one species is viable as the species' fundamental niche. In addition to viability, we also determined the metabolic by-products of growth that any one species excretes when growing on any one carbon source.

We began our community assembly process with a deliberately simple chemical environment, which supplies only one carbon source – glucose – at a constant rate. In other words, in this initial environment only one metabolic niche exists. Community assembly then consisted of multiple iterations of the following three steps (Figure 4.1):

(I) Species invasion: a species is introduced into the environment/community. Each such species is selected at random from our species pool with equal probability. The same species can be selected multiple times in the course of assembling one community.

(II) Environmental filtering: the environment serves as a filter selecting those species from the community that can persist in the environment. Specifically, we assume that those species that achieve the highest growth among all present species when consuming at least one of the available carbon sources in the environment persist. All other species go extinct. This modeling decision is based on the competitive exclusion principle, which states that species cannot coexist if they compete for the same limiting resources. We apply this filtering method because simulating the population dynamics of all species in each community is computationally infeasible, given that we simulate thousands of invasion steps in each of the thousand communities we assemble.

(III) Niche construction: all species that make up the community at a given time modify the environment by excreting by-products of metabolizing the carbon sources they use for growth.

In a first analysis, we repeated this three-step process for 5000 invasions, at which point the number of species in the communities no longer increased. In addition, we repeated the entire process of community assembly 500 times.

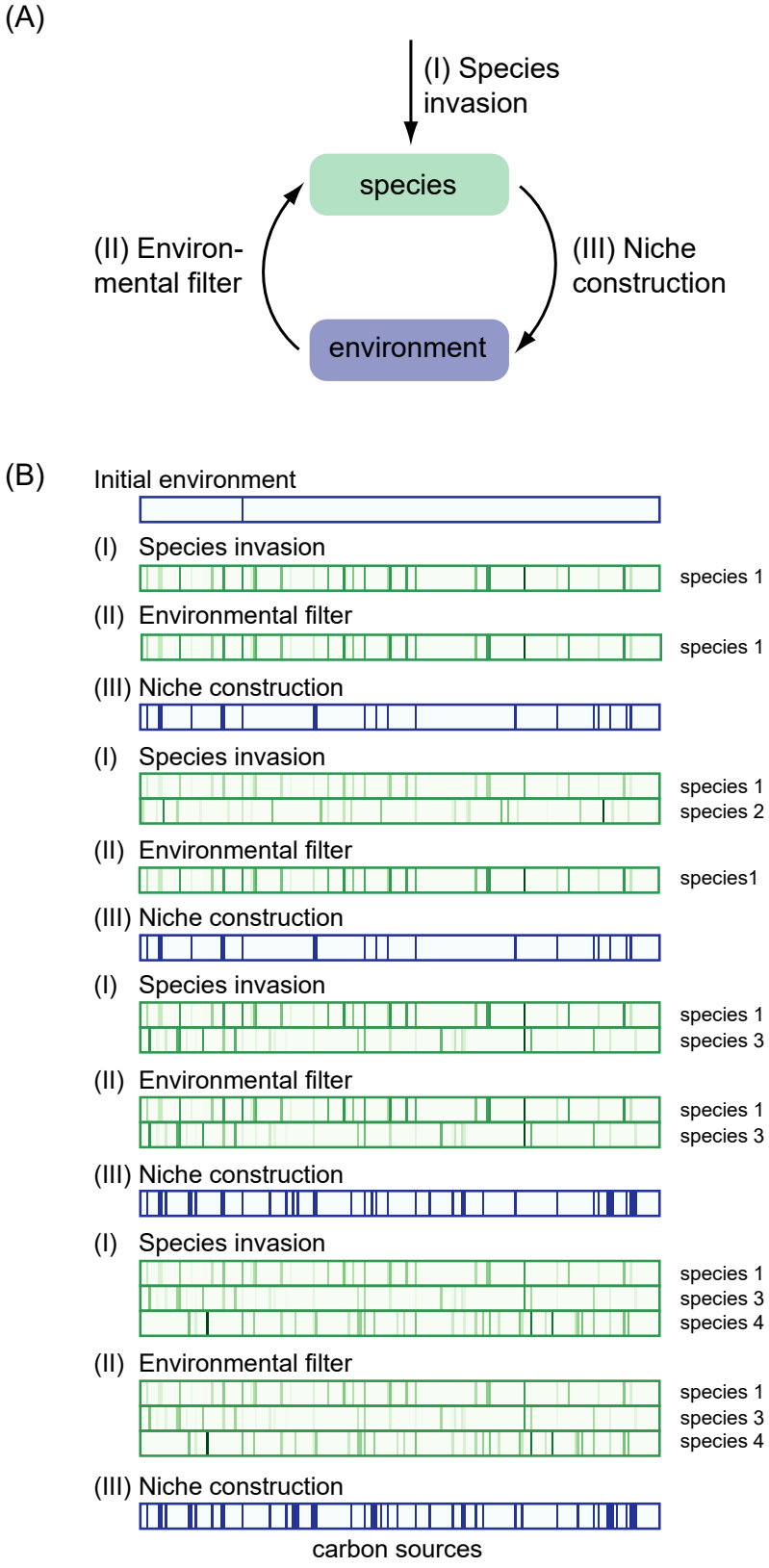

(A)
(I) Species invasion
species
(II) Environ-mental filter
(III) Niche construction
environment
(B)
Initial environment
(I) Species invasion
species 1
(II) Environmental filter
species 1
(III) Niche construction
(I) Species invasion
species 1
species 2
(II) Environmental filter
species1
(III) Niche construction
(I) Species invasion
species 1
species 3
(II) Environmental filter
species 1
species 3
(III) Niche construction
(I) Species invasion
species 1
species 3
species 4
(II) Environmental filter
species 1
species 3
species 4
(III) Niche construction
carbon sources

Figure 4.1 ***(previous page)*****: Modeling community assembly.** (A) The three steps we used to model community assembly. (I) Species (modeled as random viable metabolisms) are added at random to a standing community. (II) The environment acts as a filter, selecting successful species, i.e., those best at growing on at least one of the carbon sources available in the environment. Successful species persist in the community. (III) The species that comprise the community modify the environment through the excretion of metabolites that can serve as carbon sources for other species. (B) Example assembly trajectory of one community up to the invasion of the fourth species. Simulations begin by initializing the composition of the environment, shown as a purple rectangle at the top. The presence of each of the 223 potential carbon sources is represented as a purple vertical bar, such that each location on the horizontal x-axis corresponds to one carbon source. In this example and in most of our simulations unless otherwise stated, the initial environment contains glucose as the sole carbon source, which is indicated in the top rectangle by the single vertical purple line. After initializing the environment, we perform the first invasion with a 'species 1' chosen at random from our MCMC-derived sample of random viable networks. The species' ability to grow on each potential carbon source (whether or not the carbon source is present in the environment) is shown in the green rectangle by a green line in the x-position that corresponds to the specific carbon source. Darker green lines indicate higher growth. Species 1 grows on the only available carbon source glucose, and because no other species is yet present, we consider species 1 successful. Species 1 modifies the environment with the excretion of by-products of its growth on glucose, as shown in the second purple rectangle from the top, which are available for the second round of assembly. In this round, a randomly chose species 2 invades the community. Its growth rate on each carbon source is shown below that of species 1. Species 2 cannot grow faster than species 1 on any of the available carbon sources. It therefore goes extinct and only species 1 persists in the community. The carbon sources it excretes are available for the third assembly step. Each such step consists of a new round of (I) species invasion, (II) environmental filtering, and (III) niche construction.

Figure 4.2A-C shows the assembly dynamics for 50 randomly chosen assembled communities out of a total of 500 assembled communities, where each community had experienced up to 5000 individual species invasion events. Figure 4.2A shows that the number of persisting species (community richness, akin to alpha-diversity (Krebs, 2014)) increases rapidly until it plateaus before 5000 species invasions. A final assembled community harbors on average 35.9±0.4 species (based on 500 assembled communities). Final species richness varies somewhat, and ranges from 33 species to 38 species. Collectively, the final 500 communities harbor 64 different species, which implies there is an extensive overlap in the species making up different communities. This suggests that the order in which species invade the community have little effect on the community composition.

Figure 4.2B shows that the number of available metabolic niches, i.e., the number of carbon sources present in the environment, also increases rapidly. This increase results from the excretion of an increasing number of metabolic by-products as a community comes to host more and more species. At the same time, an increasing number of species also consumes an increasing number of carbon sources. This is why the number of free niches, that is, the number of carbon sources that are not consumed by any species in the community only rises early during community assembly (Figure 4.2C). It reaches

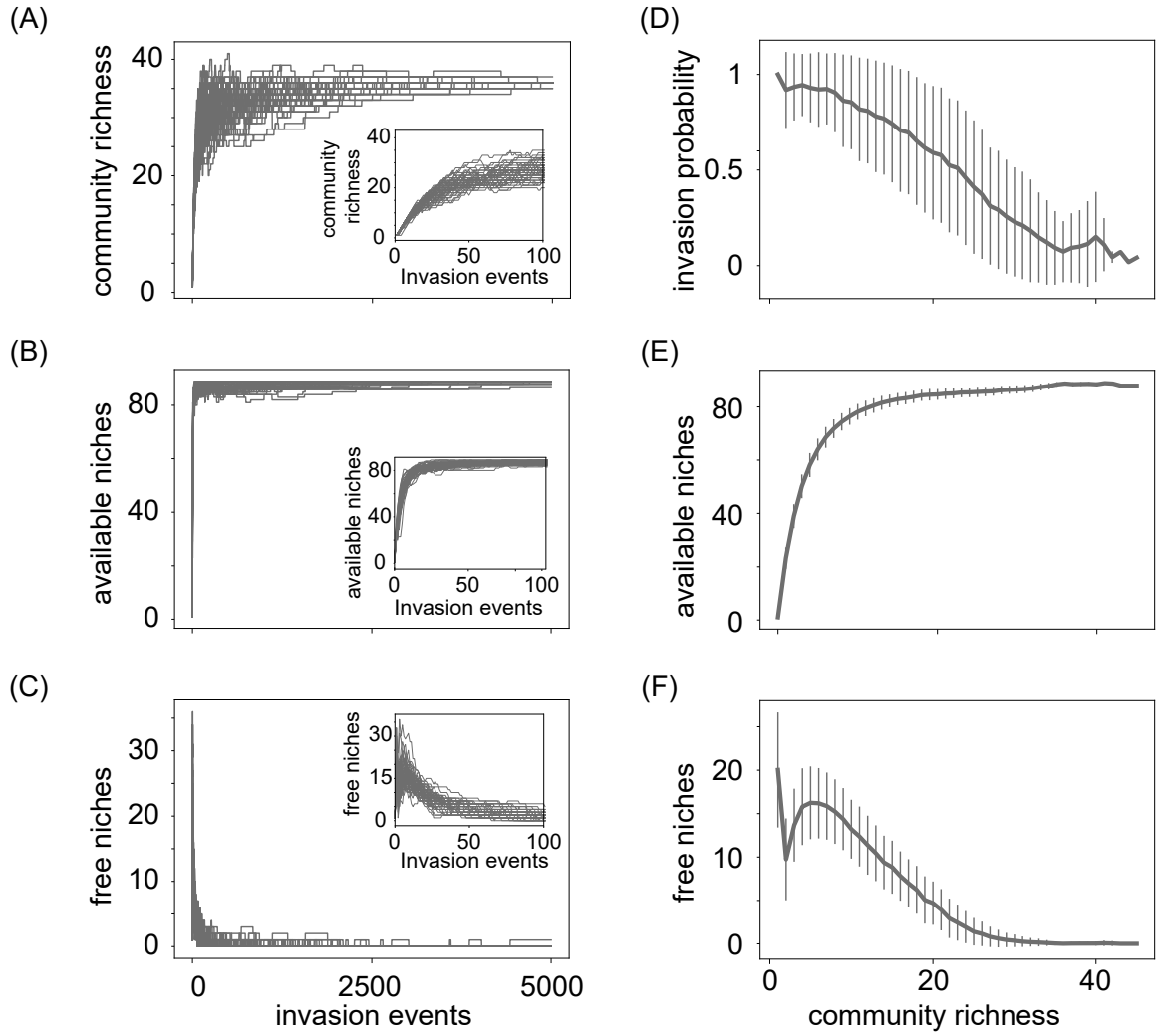


Figure 4.2: Species and niche dynamics during community assembly. (A-C) The horizontal axis shows time (number of species invasion events) during community assembly. The vertical axes show (A) community richness (number of species), (B) number of available niches (carbon sources in the environment) and (B) number of free niches (carbon sources present in the environment and not consumed by any species in the community). The insets show the same data but for up to 100 invasions. (D-F). The horizontal axis shows community richness. The vertical axes show (D) invasion probability (number of trials before a successful invasion), (E) the number of available niches (carbon sources), and (F) the number of free niches as a function of community richness. The plots show data from 50 (A-C) or 500 (D-F) assembled communities. During assembly, each community was subject to 5000 species invasion events. Error bars in plots D to F indicate one standard deviation.

a peak after approximately 6 species invasions, and then declines slowly to zero. In other words, late in the assembly process, newly invading species are more likely to consume existing carbon sources (which are excreted by resident species) than they are to lead to the excretion of new carbon sources.

We determined the probability that a new species invasion is successful, i.e., that the invading

species persists, as the number of invasion events that took place before a successful event occurred. Figure 4.2D shows that this probability declines monotonically with the number of species in a community. Figures 4.2E and F show the number of available and free niches as a function of community richness, averaged over 500 communities. The underlying data show that merely 12±2 species are needed to construct 90% of the maximal number of available niches (figure 4.2E). When a community hosts more than 6±2 species, a successful species invasion begins to create fewer niches than it fills, i.e., the number of newly excreted carbon sources becomes smaller than the additional number of carbon sources consumed. As a consequence, the number of free niches stops increasing and starts decreasing (see peak in figure 4.2F).

Taken together, the data in figure 4.2 demonstrate how important biotic interactions and in particular the construction of new metabolic niches is to the assembly process we studied. In the absence of niche construction, our environment could have only accommodated the single species that is best suited for growth on the only carbon source initially available in the environment. However, the construction of new niches does not continue throughout the assembly process. Towards the end of this process, newly invading species are more likely to fill existing niches than to help create new ones.

4.3.2 Richer initial environments lead to more diverse communities

Thus far, we had assembled communities starting from the simplest possible environment, which contains only a single carbon and energy source. We next wanted to find out how the complexity of this initial environment affects community assembly. For example, it is possible that environments containing more resources lead to the creation of more additional niches during the assembly process, which can eventually sustain richer communities. To find out, we assembled communities in environments that initially contained 1, 5, 10, 15 or 20 carbon sources. All these environments contained glucose, and we selected the remaining carbon sources at random from the list of 223 potential carbon sources (see methods).

Indeed, environments with more initial resources lead to more diverse final communities (Figure 4.3). However, initial resource differences do not translate one-to-one into final community richness. For example, an environment that initially offers 20 times more resources than the simplest environment results in communities that have only 4.0±0.1 more species (Figure 4.3). We also found that the invasion probability as a function of community richness is consistently higher for the environment that starts with the most resources, even though the difference is modest (supplementary 4.7A Fig).A higher number of available and free niches is also observed in the initially rich environment, respect to the poor one (supplementary 4.7B and C Figs).

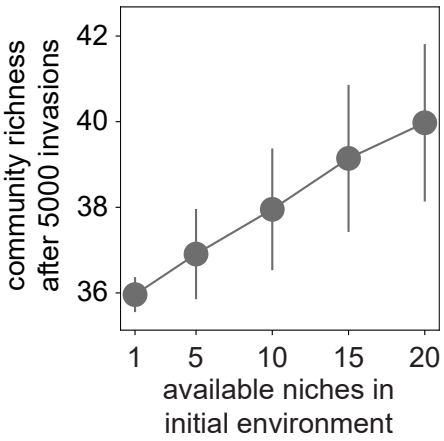


Figure 4.3: Richer initial environments lead to richer final communities. We simulated the assembly of communities in environments that offered between one and twenty carbon sources (horizontal axis) at the beginning of community assembly. The vertical axis shows the number of species (community richness) at the end of a community assembly process consisting of 5000 species invasions. Circles show mean values obtained from the assembly of 500 communities. Bars indicate one standard deviation.

4.3.3 Competition narrows the species' realized niche

Classical competition theory (Hutchinson, 1959) argues that competition is the major biological process controlling the structure of natural communities. Quantifying competition is generally difficult (Krebs, 2014) but our method allows us to quantify and manipulate competition strength in two ways. In this section, we explore how competition affects community richness. First, we can vary the strength of competition by changing the number of species that can invade the community, i.e. changing the size of the species pool. In the previous section, we had assembled communities through iterative species invasions, where we had chosen invading species from a pool of 1000 species. Now, we change the pool size to 250, 500 or 750 species. We create the species' pools at random from the original set of 1000 species.

As the size of the pool becomes larger the chances that any one species is the fastest-growing among all species on any one carbon source is lower which means that competition for resources gets stronger. In figure 4.4A we show how increasing the species pool size decreases the chances of any species from the pool of being the fastest-growing one. To calculate this, first, we computed the average number of resources on which a species showed the highest growth. This measure varies from 0 - if the species grows worse than any other species from the pool in all carbon sources - to 1 - if the species grows best than all other species in all carbon sources. We repeated this calculation for every species in a pool. Then, we calculated the average considering all species from the pool. For pools of 250, 500 and 750 species we repeated the analysis 500 times, each time with a pool created at random. In figure 4.4A we report the average of 500 repetitions.

Stronger competition, embodied in larger species pools, results in more diverse communities (fig-

ure 4.4B). This higher community richness is explained by two factors. First, as competition gets stronger, we observe a modest increase in the total number of niches available after 5000 species invasions (86.9±1.6 and 88.9±0.2 available niches for species pools of 250 and 1000 species, respectively, supplementary 4.8B Fig) while the number of free niches after 5000 invasions is not affected (all available niches are filled, supplementary 4.8C Fig).Therefore, more niches are filled. Second, we observe that stronger competition causes a reduction of the realized niche (figure 4.4C)- each persisting species consumes, on average, fewer carbon sources. (We calculated the average realized niche dividing the number of filled niches by the community richness.) Altogether, the larger community richness observed when competition is stronger results from more niches being available and filled, and from the fact that each species occupies (consumes), on average, fewer niches (resources).

A second way to vary competition strength is to alter the degree of overlap in the fundamental niches that two species occupy, i.e., the number of carbon sources that they can thrive on. Species with more similar fundamental niches show more intense competition (Bailey et al., 2013). We can manipulate this niche overlap by altering the metabolic complexity of the species that we assemble into communities, i.e., by altering the number of biochemical reactions in each metabolic network, which is possible with our sampling approach (see Methods). To this end, we created pools of metabolic networks that had 500 or 1000 more reactions than the *E. coli* model we used as a reference (iJO1366,(Orth et al., 2011)). We designated members of these pools as having metabolic complexity 'r+500' and 'r+1000', respectively. Species in these pools are on average viable on more carbon sources – they have a broader fundamental niche (supplementary 4.5 Fig). In addition, species pairs in these pools also overlap to a greater extent in their fundamental niches. We quantified this niche overlap by enumerating the number of carbon sources on which both species could grow, and divided this number by the total number of carbon sources on which at least one of the species could grow (Figure 4.4D). In sum, increased metabolic complexity increases competition among species by increasing their niche overlap.

We used species of equal metabolic complexity to assemble communities, selecting species from a pool of 1000 species (in supplementary 4.8 Fig we show the results found for pool sizes of 250, 500 and 750).

We found that communities assembled from species with high metabolic complexity are more species rich (figure 4.4E). For these three values of metabolic complexity, we observe no difference in the number of available or free niches (supplementary 4.8B and C Figs respectively). The larger community diversity observed at high metabolic complexity results from a narrower partition of niche space among the species (figure 4.4F).

We also explored the consequences of reducing metabolic complexity below that of r. We assembled communities with species that contain 500 or 1000 fewer reactions than the reference model (pools designated 'r-500' and 'r-1000'). The results found for metabolic complexity lower than r are more

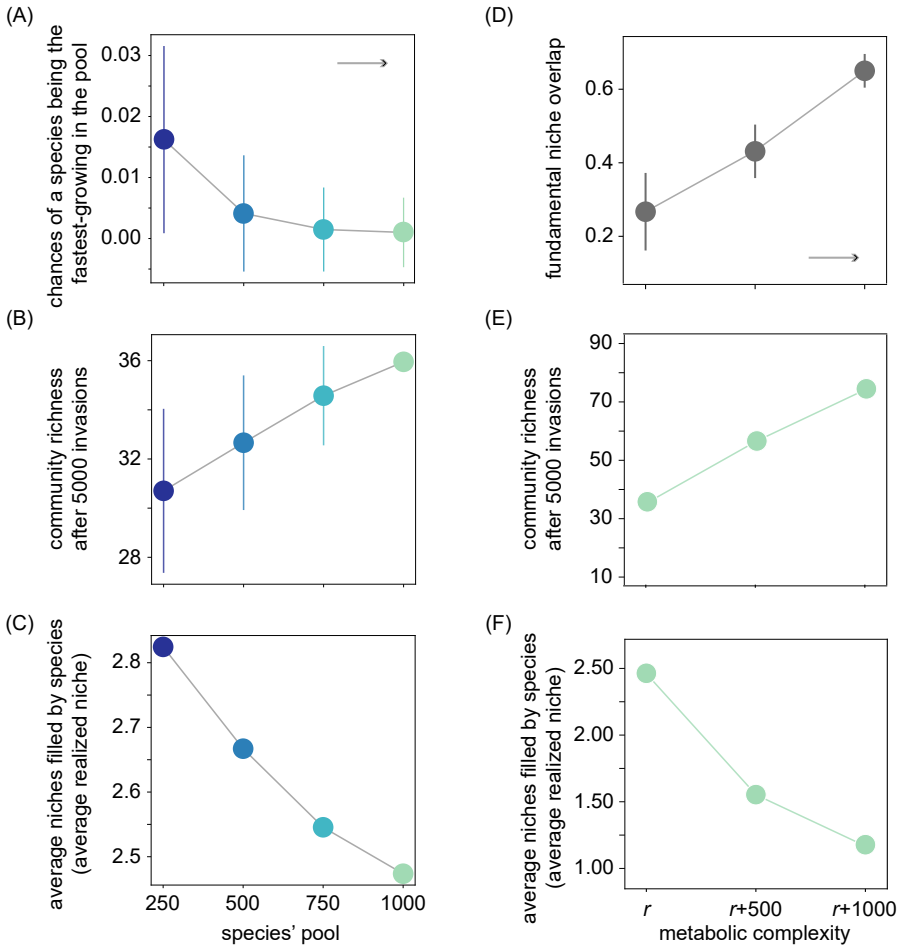


Figure 4.4: Competition causes a narrower division of the available niches. (A-C) The horizontal axis shows the number of species in the species pool from which communities are assembled. Larger species pools imply higher competition among species. Vertical axes show (A) the probability of a species to be the fastest growing on a carbon source, average across carbon sources and across the whole pool, (B) community richness (number of species), and (C) average number of niches filled by each species (average realized niche, quantified as the number of available niches minus the number of free niches and divided by the total number of species in a community). (D-F) The horizontal axis shows the metabolic complexity of the species used for community assembly, i.e., their total number of biochemical reactions relative to the reference r=2583 of *E. coli* model iJO1366. (D) The vertical axis shows the average overlap in fundamental niche between two species pairs, quantified as the number of carbon sources on which both species can grow, divided by the union of carbon sources on which either species can grow. Higher metabolic complexity resulted in higher niche overlap between species, which increases the potential for competitive interactions between species. (E) community richness, and (F) average realized niche breadth (quantified as in B). Arrows in A and D indicate the direction in which competition gets stronger. Circles in B, C, E and F show averages over 500 assembled communities after simulating 5000 invasion events per community. Circle colors indicate the size of the species pool used to assemble the communities (a color change from blue to green corresponds to an increase in the pool size from 250 to 1000 species). Bars show one standard deviation.

complex than those observed for metabolic complexity higher than r. As in the analysis where we changed the species' pool, here it is the combination of two factors that explains the community

richness, the total number of niches filled and the average realized niche.

In sum, we observe that when competition is strong that results in a narrower division of the niche space each species occupies. Everything else being equal, this derives in more diverse communities.

4.4 Discussion

The 'ecological limits' and 'diversity begets diversity' models were proposed to explain the forces that shape biodiversity on Earth. Therefore, it is reasonable that some analyses performed to explore the validity of these hypotheses (Brockhurst et al., 2007; Gómez and Buckling, 2013; Jousset et al., 2016) were done considering both evolutionary and ecological processes acting together, just as it happens in nature. However, we believe that to test whether diversity inhibits or promotes further diversity we should disentangle purely evolutionary from ecological processes. With this in mind, here we simulate the assembly of thousands of communities were we control the amount of new species allowed to invade a community and monitor the community's response to the invasion.

Our results suggest that the dichotomy between diversity begets diversity and ecological limits is a false dichotomy and that both principles may be correct. At initial stages of community assembly species-species interactions, in the form of cross-feeding and competition, facilitate diversity. Species create niches through their metabolic excretions opening the opportunity for new species to invade as they thrive by consuming those excretions (figure 4.2).Competition causes a narrower partition of niche space which can result in larger diversity (figure 4.4). At later stages in the assembly process, no new niches can be created, niche space cannot be partitioned finer, successful species invasions become rare and diversity reaches a ceiling. In sum, community diversity has limits, but those are not just externally imposed but are modified by the species themselves.

Two other conclusions emerge from our results. First, that observing that all niches are filled does not necessary mean that the diversity has reached its ceiling. Competition can raise community richness even when all niches are filled by narrowing the niche space of species (supplementary 4.6 Fig). Second, observing a slowdown in speciation or diversification rate is not proof of ecological limits model. We quantified the probability of successful invasions and found that this probability decreases as the community richness increases (figure 4.2D and supplementary 4.7 and 4.9 Figs).These results suggest that we expect to observe a diversification slowdown as community diversity gets larger, in agreement with the pervasive observation that diversification slows down as evolution proceeds (Condamine et al., 2019). And this decrease is observed even in the range of community richness where niches are being created by the species, that is, in ranges where 'diversity begets diversity' applies.

Competition is commonly thought to stimulate the diversification of species (Dieckmann and Doebeli, 1999a; Bailey et al., 2013; Meyer and Kassen, 2007; Tilman, 1982; Begon et al., 2006; Johnson et al., 2012), although it can also prevent it (Bailey et al., 2013; Brockhurst et al., 2007; Gómez and

Buckling, 2013). In our work, a successful invasion increases competition as the available resources have to be shared between more species. After any one invasion, communities show an increase, no change, or a net decrease in diversity if multiple species go extinct as a result of the invasion (Figure 4.2A).These disparate results found after a species invasion are in agreement with the contrasting results found in the literature. However, in our simulations, we observe that by the end of the assembly process (once community richness no longer increases) those communities that experienced stronger competition show larger species diversity (figure 4.4) and we conclude that, overall, competition fosters diversity.

In an organism, traits might be interconnected. When a trait is advantageous for one function but is associated to another trait that results somehow disadvantageous, trade-offs arise. Trade-offs are thought to be important for determining community structure (Litchman and Klausmeier, 2008; Tilman, 1982; Begon et al., 2006; Johnson et al., 2012). And our results indirectly suggest that trade-offs in resource use are important for the communities assembled here. If there were no such trade-offs, we would expect that one or a few species consumed all available resources. Instead, the richness of species found in the communities assembled here suggests that trade-offs in the use of resources are present. It is important to point out that these trade-offs arise simply from the set of reactions included in the metabolic networks.

Our work has several limitations. First, we study a very constrained subspace of Hutchinsonian niche space-the multidimensional space considering all environmental factors required for the survival of a species (Hutchinson, 1957). Specifically, we only consider the sources of carbon and energy that sustain heterotrophic organisms because they are frequently cross-fed in between bacteria (D'Souza et al., 2018). We also only consider a small fraction of the thousands of conceivable natural carbon sources.

Second, the overlap in the set of carbon sources on which species grow (the similarity in their fundamental niche) may be a consequence of the method we used to create metabolic networks for community assembly. Specifically, we constrained these networks to be viable on glucose, which may have increased their similarity in terms of the carbon sources on which they are viable, because these networks must share subset of reactions required for growth on glucose (Barve et al., 2012). To explore how this choice may have affected our analysis, we created random metabolic networks required to be viable on carbon sources other than glucose (acetate, pyruvate, serine, alanine and lactose) (supplementary 4.10 Fig). Communities assembled using these networks differed little in community diversity. This observation holds true when communities are assembled mixing species whose metabolic networks were selected to be viable in one carbon source from a pair, suggesting that this limitation does not affect our observations dramatically.

Third, we observed that communities can be composed of up to 80 species, which depended, directly

or indirectly, on the species consuming the initial resource glucose. Unfortunately, our current modeling approach cannot answer how abundant these species are, and how much glucose would be required to sustain the community. Understanding the relative abundances of the species in communities like those we model thus remains an important task for future work.

Though outside the scope of our work, the modeling framework used here can be used to study, for example, successional dynamics (Nemergut et al., 2007; Dini-Andreote et al., 2014; Lockwood et al., 1997; Chase, 2003) and the role of historical contingency (or priority effects) in the assembly process (Chase, 2003; Fukami, 2015). What is more, whole-community genome sequencing, together with semi-automatic methods for metabolic reconstruction, facilitate the creation of genome-scale metabolic models for not just one organism, but for multiple organisms in a community (DeJongh et al., 2007; Wang et al., 2018; Dias et al., 2015).Thus, it may soon be possible to conduct an analysis like ours with metabolic networks characterized in a community from the wild. While such an analysis cannot control all the variables that our computational work can control, it can go beyond a proof of principle, explain actual limits on community diversity, and identify rules of community assembly important in nature. It will be especially suited for the analysis of those complex communities where cross-feeding interactions are important, such as that of the human gut (Magnúsdóttir et al., 2017), the soil (Baran et al., 2015), or planktonic organisms (Enke et al., 2019).

4.5 Methods

4.5.1 Flux balance analysis (FBA)

Flux balance analysis (FBA) is a computational method to predict metabolic fluxes – the rate at which chemical reactions convert substrates into products – of all biochemical reactions in the metabolism of an organism (Orth et al., 2010). Because these reactions form a complex chemical reaction network, we refer to them also as a metabolic network. Metabolic networks have been characterized for hundreds of species (King et al., 2016), and are often also referred to as genome-scale metabolic models of these species.

FBA requires information about the stoichiometry of all chemical reactions in a metabolic network. It makes two central assumptions. The first is that cells (metabolisms) are in a metabolic steady-state. The second is that cells effectively optimize some metabolic property such as biomass growth. Additional constraints can be incorporated into the optimization problem that FBA solves, in order to account for the thermodynamic and enzymatic properties of a network's biochemical reaction (Orth et al., 2010). The optimization problem that FBA solves can be formalized as a linear programming problem (Orth et al., 2010; Varma and Palsson, 1994b) in the following way:

$$\begin{aligned} &Maximize\ v_{growth} \\ &s.t.\ Sv = 0 \\ &l_i \leq v_i \leq u_i \end{aligned} \tag{4.1}$$

Here, S is the stoichiometric matrix, a matrix of size $m \times r$ that mathematically describes the stoichiometry of the network's metabolic reactions. The integer m denotes the number of metabolites, and the integerr denotes the number of reactions in the network. Each entry S_{ij} of the stoichiometric matrix contains the stoichiometric coefficient with which metabolite i participates in reaction j. The vector v (of size r) harbors the metabolic flux through each reaction in the network. v_{growth} specifies the flux through the biomass growth reaction. Fluxes through biochemical reactions are restricted by lower and upper bounds that constrain the flux through each reaction in the network. These bounds are given by the variables l and u, respectively, which are vectors of size r. We performed FBA optimization with the GNU Linear Programming Kit (GLPK; http://www.gnu.org/software/glpk).

4.5.2 Modeling species with random viable networks

To sample random viable networks, we began with a specification of desired metabolic characteristics. For example, for many of our analyses, we wanted to create networks that are at least viable on glucose as a sole carbon source, and that comprise as many reactions as *E. coli* (i.e., 2583 reactions in the widely used *E. coli* metabolic model iJO1366). To identify a single initial network fulfilling these requirements, we first performed Flux Balance Analysis (see previous section) on the pan-metabolic network of 7222 reactions (Barve and Wagner, 2013; San Roman and Wagner, 2018) in a chemically minimal environment with glucose as the only source of carbon. Of all reactions in this pan-metabolic network, 1263 reactions showed non-zero metabolic flux. We included all these reactions in the initial network, which ensured the network's viability on glucose. We then chose the remaining (1320) reactions needed to arrive at an equal number of reactions as the *E. coli* metabolism at random (with a uniform distribution) from the pan-metabolic network.

This initial network was the starting point of MCMC sampling from the space of all possible metabolic networks. The MCMC procedure performs a long random walk through this space. Each step in this walk consists of a reaction swap, in which a reaction from the current network is randomly chosen for deletion, while a randomly chosen reaction from pan-metabolism but absent in the current network is added to the current network. If the network thus modified remains viable after this reaction swap, the swap is accepted, and the network is modified with a further reaction swap. In contrast, if the reaction swap disrupts viability on the desired carbon source(s), such as glucose, the swap is rejected and a new swap is tried. Modifying metabolic networks through reaction swaps ensures that the number of reactions in the network remains constant and equal to the number of reactions in the

initial network.

This MCMC procedure creates a long sequence of metabolic networks, all of them viable on the desired carbon sources. As the number of reaction swaps increases, the number of reactions that the altered network shares with the initial network becomes smaller and smaller, until its complement of reactions becomes effectively randomized, which occurs after approximately 5000 successful swaps (Samal et al., 2010). We stored such a randomized network for further analysis after 5000 successful swaps, and repeated this procedure from the initial network to create 1000 random networks viable on specific carbon sources, such as glucose. During this process we did not alter the exchange reactions of the starting network, which ensures that in the randomized networks the same metabolites can be exchanged with the environment as in *E. coli*. Furthermore, we used the biomass reaction from *E. coli* iJO1366 to assess the viability of networks during MCMC sampling. We note that networks required to be viable on specific carbon sources such as glucose also often happen to be viable on multiple additional carbon sources, a consequence of the reticulate nature of metabolisms (Barve and Wagner, 2013).

4.5.3 Potential carbon sources

As in (Hutchinson, 1959), we performed FBA to examine all metabolites that the pan-metabolic network could use as sole carbon sources. To this end, we used the biomass reaction of the model of *E. coli* iJO1366 in the pan-metabolic network. We performed FBA in a simulated environment in which one candidate carbon source was available at a time and assumed that ammonium, calcium, chloride, cobalt, copper, iron, magnesium, manganese, molybdate, nickel, oxygen, phosphate, potassium, protons, sodium, sulphate and zinc were available in the environment in non-limiting amounts. We considered a metabolite a 'candidate' or 'potential' carbon source if it contained at least one carbon atom, and if an exchange reaction for this metabolite was present in the network of *E. coli*. In this way, we identified 223 potential carbon sources. Any metabolism sampled from the pan-metabolic network will be viable on a subset of these carbon sources.

4.5.4 Identifying species traits with Flux Balance Analysis (FBA)

For every random viable metabolic network ('species') we considered, we used FBA to determine its viability on all 223 potential carbon sources. To this end, we again performed FBA in a minimal environment where only one potential carbon source was available at a time, as explained in the previous section. The result was a predicted maximal biomass growth flux for each species on each of the 223 potential carbon sources. We used these predicted growth rates in the second step of our simulation of community assembly, i.e., in the environmental filter which accounts for competition among species, where the environment selects those species that grow best on the available resources.

We also used FBA to analyze the niche construction potential of every species. We were interested in those metabolites that could be excreted by one species, and thus potentially serve as carbon sources for some other species. Therefore, we identified those metabolites from the list of potential carbon sources that could be produced as by-product of growth when a species consumes a given carbon source. To this end, we allowed consumption of this carbon source (with a maximal uptake rate of 10 $mmol\, gDW^{-1}\, h^{-1}$) in an otherwise minimal environment, and maximized the production of each potential carbon source, while constraining the species' biomass production to be greater than zero. If a potential carbon source could be produced at a rate greater than zero under this constraint, we considered that the carbon source was excreted into the environment.

The results of FBA indicate that when microorganisms grow at their maximum, no potential carbon source is excreted. All carbon consumed is used for biomass production. However, we know from experiments that metabolites leak through the cell membrane and are excreted as result of overflow metabolism (Paczia et al., 2012). Here, we assume that all species experience the same biomass reduction by losing metabolites to the environment.

4.6 Supplementary material

4.6.1 Two factors affect species diversity when metabolic complexity is reduced below r

As observed in figure 4.4A, increasing metabolic complexity from r to $r + 1000$ increases fundamental niche overlap between species. The reason is that larger metabolic networks contain more reactions from the pan-metabolic network (whose number of reactions is fixed at 7222 reactions), such that a larger fraction of reactions is expected to be identical between pairs of random viable networks. This larger overlap of reactions increases the likelihood that two networks are viable on the same carbon sources, which causes a greater overlap in the fundamental niche. Conversely, when we decrease metabolic complexity below r, we expect that at some value fundamental niche overlap will reach a minimum, below which it will start to increase again. The reason is that any two random networks viable on glucose must share reactions essential for viability on glucose. As metabolic complexity decreases, more reactions in each network will be essential, and any two networks will share a greater fraction of reactions. This will also increase the overlap between pairs of networks in terms of the carbon sources on which both networks grow. In other words, it will increase the overlap in their fundamental niche, a proxy for the strength of competition.

To find out whether this is the case, we created pools of metabolic networks that had 500 or 1000 fewer reactions than the *E. coli* model we used as a reference. We refer to these species pools as the 'r-500' and 'r-1000' pools. Coincidentally, metabolic complexity r lies close to the point of

minimal fundamental niche overlap (supplementary 4.8A Fig). Thus, competition strength increases if we increase or decrease metabolic complexity from r.

We then used species of metabolic complexity $r - 500$ and $r - 1000$ to assemble communities. The community richness after 5000 invasion events shows a complex pattern (r, $r - 500$ and $r - 1000$) that can be explained by the same factors that explain community richness when we changed species pool size (Figure 4.4A-C). As we reduce metabolic complexity below r, competition becomes stronger, which results in a division of niche space into narrower realized niches (supplementary 4.8F Fig). Although one might expect that communities get more diverse with narrower individual niches, this is not what we observe. The reason is that the number of filled niches also influences community diversity at low metabolic complexity. Specifically, as metabolic complexity decreases below r, fewer niches become available, and more niches are left unfilled (supplementary 4.8B and C Figs). In sum, at low metabolic complexity, both the breadth of realized niches and the total number of niches filled are needed to explain community diversity. In contrast, at high metabolic complexity, as discussed in the main text and shown in Figures 4.4D-F and supplementary 4.8 Fig, only the breadth of realized niches determines community diversity.

4.6.2 Supplementary figures

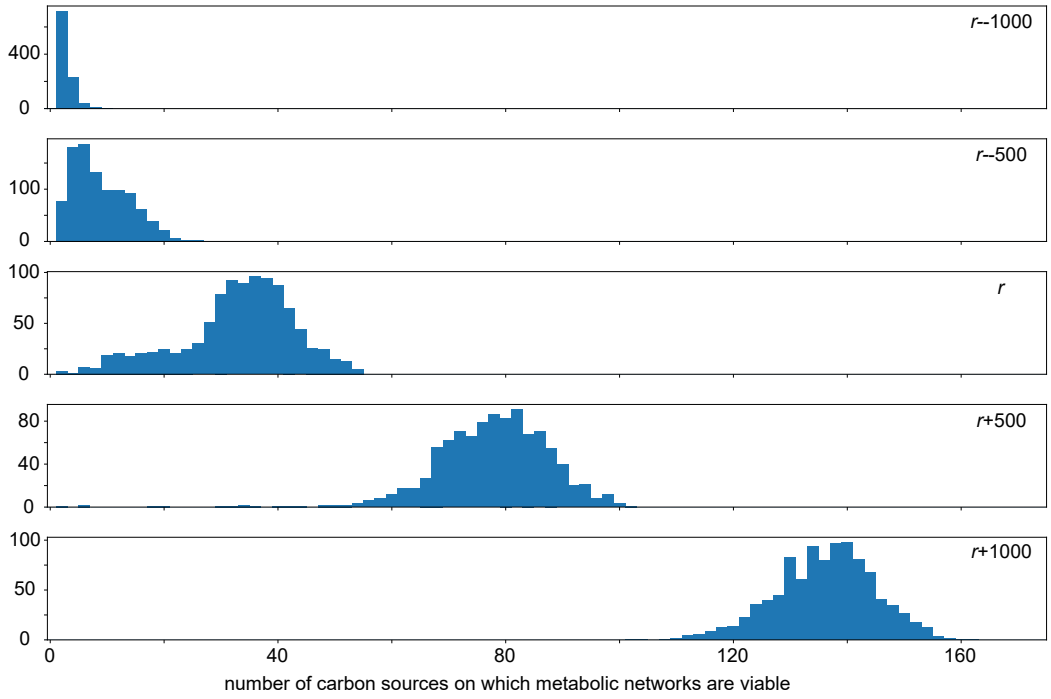


Figure 4.5: For species whose metabolic complexity increases from top ($r - 1000$ reactions) to bottom ($r + 1000$ reactions), the histograms show the fundamental niche width, i.e., the number of carbon sources on which the metabolic networks representing these species are viable. The variable r indicates the 2583 reactions of the reference metabolism of *E. coli* (Orth et al., 2011). Each panel shows data from 1000 random viable metabolic networks (species).

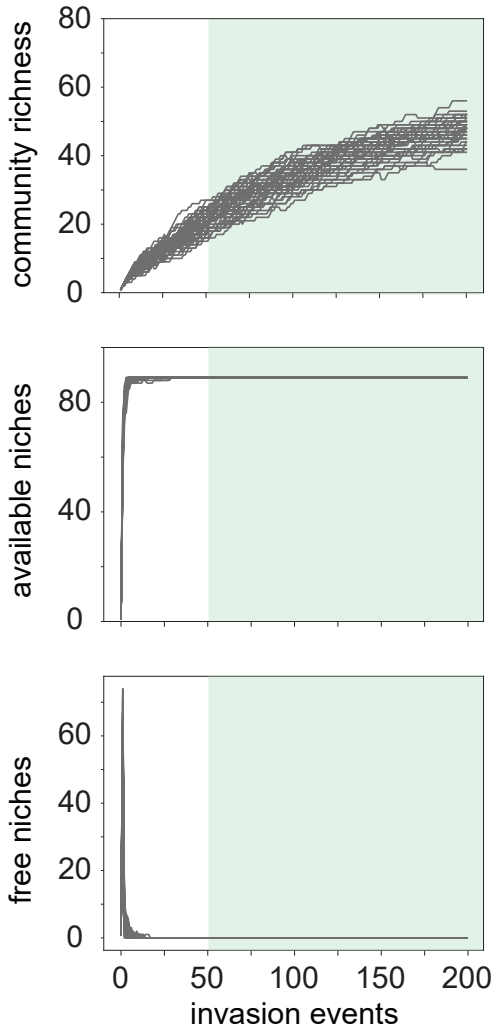


Figure 4.6: Competition for resources contributes to increase the diversity of communities even in the absence of free niches. We show the results of simulating the assembly of 50 communities made up of species modeled with networks of high metabolic complexity (which have 1000 reactions more than the metabolic reference network of *E. coli*, $r + 1000$). Consecutive invasion events (horizontal axis) change (A) community richness, (B) available niches, and (C) free niches. The area shaded in green highlights how community richness keeps increasing even when no more niches are being created and no niches remain free, as the available niches are divided more tightly between the species.

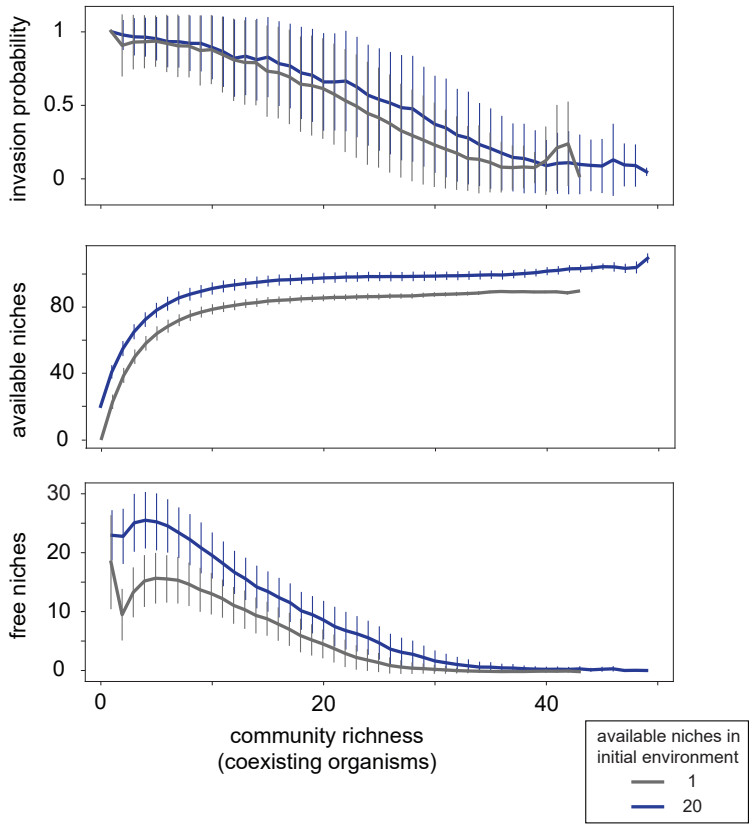


Figure 4.7: The vertical axes show (A) invasion probability, (B) the number of available niches (carbon sources), and (C) the number of free niches, plotted against community richness (horizontal axis). These are results of assembly in an environment that initially contains only glucose (grey line) or glucose plus 19 other carbon sources selected at random from a list of 223 potential carbon sources (blue line). Results show mean values obtained from the assembly of 500 communities. Vertical bars indicate one standard deviation.

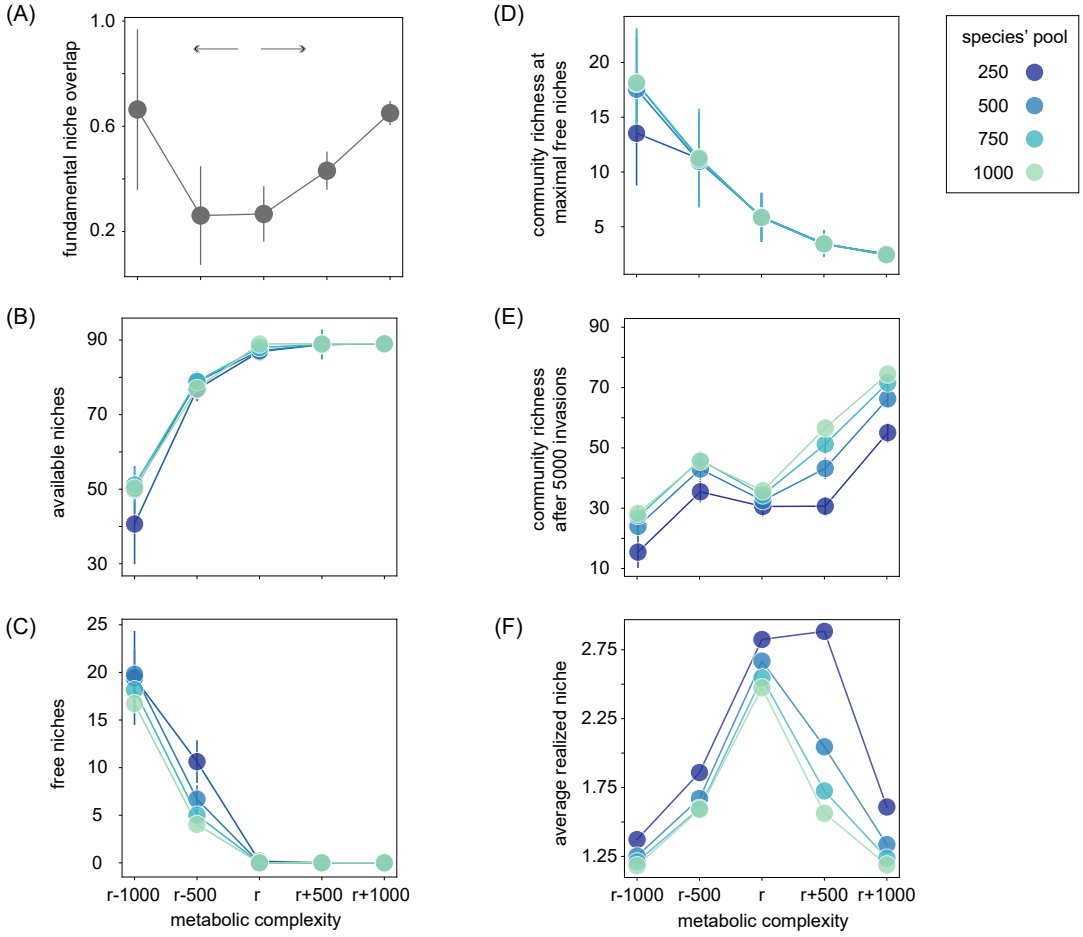


Figure 4.8: The horizontal axis shows the metabolic complexity of the species used for community assembly, i.e., their total number of biochemical reactions relative to the reference *E. coli* model iJO1366 (r). The vertical axis shows (A) the average overlap in fundamental niche between two species pairs, quantified as the number of carbon sources on which both species can grow, divided by the union of carbon sources on which either species can grow. Higher metabolic complexity resulted in higher niche overlap between species, which increases the potential for competitive interactions between species. Arrows indicate the direction in which competition gets stronger. (B) and (C) the available and free niches, respectively, after 5000 invasion events. (D) community richness at maximal number of free niches during assembly,(E) community richness, and (F) average realized niche breadth (quantified as in figure 4.4B). Circles in (B)-(F) show averages over 500 assembled communities after simulating 5000 invasion events per community. Circle colors indicate the size of the species pool used to assemble the communities (a color change from blue to green corresponds to an increase in the pool size from 250 to 1000 species). Bars show one standard deviation.

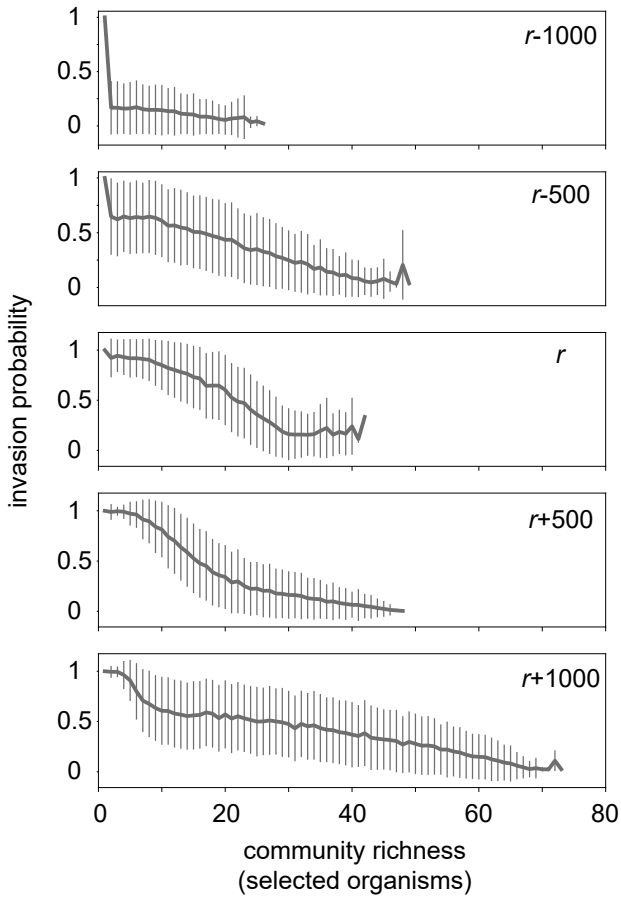


Figure 4.9: Probability of successful species invasion(vertical axes) as a function of community richness (horizontal axes) for networks of varying complexity (number of reactions relative to the reference r=2583 reactions of *E. coli* model iJO1366 (Orth et al., 2011), see inset text) . Results are based on mean values calculated from the assembly of 50 communities. Bars indicate one standard deviation.

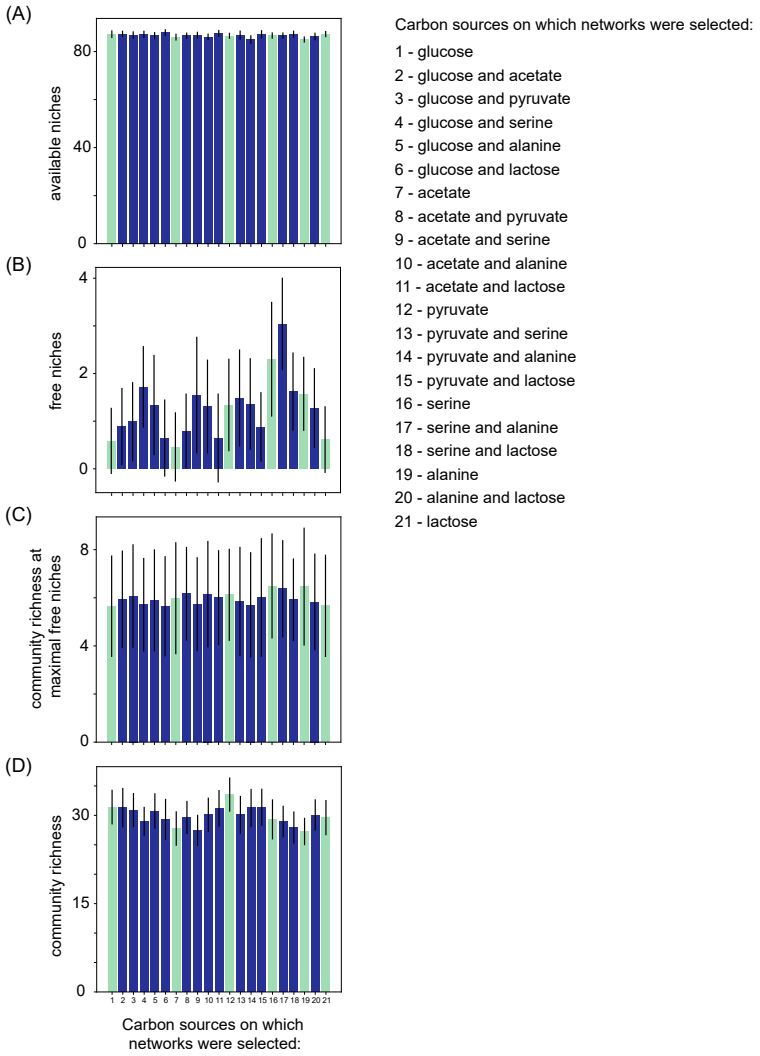


Figure 4.10: In this figure we show how various community properties change when the species that make up those communities are required to be viable on carbon sources other than glucose. We used MCMC-sampling to create metabolic networks viable on one of the following carbon sources: lactose, acetate, pyruvate, alanine and serine. Specifically, we created 300 networks that are viable on each carbon source, and that contained as many reactions (2583) as the reference network of *E. coli* (iJO1366, (Orth et al., 2011)). We used these networks to assemble communities where all species in the community were required to be viable on the same carbon source (green bars). In addition, we assembled communities from species whose networks were required to be viable on either carbon source of a pair of carbon sources (blue bars, see figure legend). Specifically, we created a mixed set of 300 species, where 150 species were required to be viable on one of the carbon sources, and the other 150 species were required to be viable on the other source. From this mixed set, we chose at random 150 species which constituted the mixed pool. By doing so, the mixed pool can be formed with 150 species viable on one of the carbon sources, 150 species viable on the other carbon source or any intermediate composition. We used this mixed pool to assemble 150 communities. Before simulating the assembly of each community, we selected a new mixed pool. We simulated the assembly of communities through 5000 species invasions. Horizontal axes show the carbon source or carbon source pair on which species in the assembly pool were required to be viable. Vertical axes show mean (bar) and standard deviation (whisker) of the (A) number of final available niches, (B) number of final free niches, (C) community richness at maximal number of free niches during assembly, and (D) final community richness (D). The data show that the carbon source(s) on which the metabolic networks were required to be viable have very little influence on community properties.

Bibliography

M. AbuOun, P. F. Suthers, G. I. Jones, B. R. Carter, M. P. Saunders, C. D. Maranas, M. J. Woodward, and M. F. Anjum. Genome Scale Reconstruction of a Salmonella Metabolic Model: Comparison of similarity and differences with a commensal Escherichia coli strain. *Journal of Biological Chemistry*, 284(43):29480–29488, Oct. 2009. doi: 10.1074/jbc.M109.005868. URL http://www.jbc.org/content/284/43/29480.abstract.

R. Albert. Scale-free networks in cell biology. *Journal of Cell Science*, 118(21):4947, Nov. 2005. doi: 10.1242/jcs.02714. URL http://jcs.biologists.org/content/118/21/4947.abstract.

B. Allen, J. Gore, and M. A. Nowak. Spatial dilemmas of diffusible public goods. *eLife*, 2:e01169, Dec. 2013. ISSN 2050-084X. doi: 10.7554/eLife.01169. URL https://elifesciences.org/articles/01169.

S. D. Allison. A trait-based approach for modelling microbial litter decomposition. *Ecology Letters*, 15 (9):1058–1070, Sept. 2012. ISSN 1461-023X. doi: 10.1111/j.1461-0248.2012.01807.x. URL https://doi.org/10.1111/j.1461-0248.2012.01807.x.

E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, and A.-L. Barabási. Global organization of metabolic fluxes in the bacterium Escherichia coli. *Nature*, 427(6977):839–843, Feb. 2004. ISSN 1476-4687. doi: 10.1038/nature02289. URL https://doi.org/10.1038/nature02289.

L. Andersson. The driving force: species concepts and ecology. *TAXON*, 39(3):375–382, Aug. 1990. ISSN 0040-0262, 1996-8175. doi: 10.2307/1223084. URL https://onlinelibrary.wiley.com/doi/abs/10.2307/1223084.

A. Astorga, J. Oksanen, M. Luoto, J. Soininen, R. Virtanen, and T. Muotka. Distance decay of similarity in freshwater communities: do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography*, 21(3):365–375, Mar. 2012. ISSN 1466-822X. doi: 10.1111/j.1466-8238.2011.00681.x. URL https://doi.org/10.1111/j.1466-8238.2011.00681.x.

K. C. Atwood, L. K. Schneider, and F. J. Ryan. Periodic Selection in Escherichia Coli. *Proceedings*

of the National Academy of Sciences, 37(3):146, Mar. 1951. doi: 10.1073/pnas.37.3.146. URL http://www.pnas.org/content/37/3/146.abstract.

S. F. Bailey, J. R. Dettman, P. B. Rainey, and R. Kassen. Competition both drives and impedes diversification in a model adaptive radiation. *Proceedings of the Royal Society B: Biological Sciences*, 280(1766):20131253, Sept. 2013. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2013.1253. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2013.1253.

A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry*, 50(21):4402–4410, 2011. doi: 10.1021/bi2002289. URL http://dx.doi.org/10.1021/bi2002289.

R. Baran, E. L. Brodie, J. Mayberry-Lewis, E. Hummel, U. N. Da Rocha, R. Chakraborty, B. P. Bowen, U. Karaoz, H. Cadillo-Quiroz, F. Garcia-Pichel, and T. R. Northen. Exometabolite niche partitioning among sympatric soil bacteria. *Nature Communications*, 6(1):8289, Sept. 2015. ISSN 2041-1723. doi: 10.1038/ncomms9289. URL https://doi.org/10.1038/ncomms9289.

M. Barluenga, K. N. Stölting, W. Salzburger, M. Muschick, and A. Meyer. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, 439:719, Feb. 2006. URL http://dx.doi.org/10.1038/nature04325.

A. Barve and A. Wagner. A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature*, 2013.

A. Barve, J. F. M. Rodrigues, and A. Wagner. Superessential reactions in metabolic networks. *Proceedings of the National Academy of Sciences*, 109(18):6810, May 2012. URL http://www.pnas.org/content/109/18/E1121/1.abstract.

A. Barve, S.-R. Hosseini, O. C. Martin, and A. Wagner. Historical contingency and the gradual evolution of metabolic properties in central carbon and genome-scale metabolisms. *BMC Systems Biology*, 8(1):48, Apr. 2014. ISSN 1752-0509. doi: 10.1186/1752-0509-8-48. URL https://doi.org/10.1186/1752-0509-8-48.

D. A. Beard, S.-d. Liang, and H. Qian. Energy balance for analysis of complex metabolic networks. *Biophysical Journal*, 83(1):79–86, July 2002. ISSN 0006-3495. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1302128/.

M. Begon, C. R. Townsend, and J. L. Harper. *Ecology: from individuals to ecosystems*. Blackwell Pub, Malden, MA, 4th ed edition, 2006. ISBN 978-1-4051-1117-1.

S. T. Behmer and A. Joern. Coexisting generalist herbivores occupy unique nutritional feeding niches. *Proceedings of the National Academy of Sciences*, 105(6):1977–1982, Feb. 2008. URL http://www.pnas.org/content/105/6/1977.abstract.

A. Betz and B. Chance. Phase relationship of glycolytic intermediates in yeast cells with oscillatory metabolic control. *Archives of Biochemistry and Biophysics*, 109(3):585–594, Mar. 1965. ISSN 0003-9861. doi: 10.1016/0003-9861(65)90404-2. URL http://www.sciencedirect.com/science/article/pii/0003986165904042.

S. Blasche, Y. Kim, R. Mars, E. Kafkia, M. Maansson, D. Machado, B. Teusink, J. Nielsen, V. Benes, R. Neves, U. Sauer, and K. R. Patil. Emergence of stable coexistence in a complex microbial community through metabolic cooperation and spatio-temporal niche partitioning. preprint, Microbiology, Feb. 2019. URL http://biorxiv.org/lookup/doi/10.1101/541870.

Z. D. Blount, R. E. Lenski, and J. B. Losos. Contingency and determinism in evolution: Replaying life's tape. *Science*, 362(6415):eaam5979, Nov. 2018. doi: 10.1126/science.aam5979. URL http://science.sciencemag.org/content/362/6415/eaam5979.abstract.

J. A. Bondy and U. S. R. Murty. *Graph theory with applications*. North Holland, New York, 1976. ISBN 978-0-444-19451-0.

M. A. Brockhurst, N. Colegrave, D. J. Hodgson, and A. Buckling. Niche Occupation Limits Adaptive Radiation in Experimental Microcosms. *PLOS ONE*, 2(2):e193, Feb. 2007. doi: 10.1371/journal.pone.0000193. URL https://doi.org/10.1371/journal.pone.0000193.

A. Broido and A. Clauset. Scale-free networks are rare. *Nature Communications*, 10, Jan. 2018. doi: 10.1038/s41467-019-08746-5.

S. Brooks. Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):69–100, 1998. ISSN 1467-9884. doi: 10.1111/1467-9884.00117. URL http:https://doi.org/10.1111/1467-9884.00117.

A. P. Burgard, E. V. Nikolaev, C. H. Schilling, and C. D. Maranas. Flux Coupling Analysis of Genome-Scale Metabolic Network Reconstructions. *Genome Research*, 2004. doi: 10.1101/gr.1926504.

V. Calcagno, P. Jarne, M. Loreau, N. Mouquet, and P. David. Diversity spurs diversification in ecological communities. *Nature Communications*, 8:15810, June 2017. URL https://doi.org/10.1038/ncomms15810.

B. J. Callahan, T. Fukami, and D. S. Fisher. Rapid evolution of adaptive niche construction in experimental microbial populations. *Evolution*, 68(11):3307–3316, 2014. ISSN 1558-5646. doi: 10.1111/evo.12512. URL http://dx.doi.org/10.1111/evo.12512.

R. Carlson, D. Fell, and F. Srienc. Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnology and Bioengineering*, 79(2):121–134, July 2002. ISSN 0006-3592, 1097-0290. doi: 10.1002/bit.10305. URL http://doi.wiley.com/10.1002/bit.10305.

P. Cermeño, I. G. Teixeira, M. Branco, F. G. Figueiras, and E. Marañón. Sampling the limits of species richness in marine phytoplankton communities. *Journal of Plankton Research*, 36(4):1135–1139, May 2014. ISSN 0142-7873. doi: 10.1093/plankt/fbu033. URL https://doi.org/10.1093/plankt/fbu033.

S. Chaffron, H. Rehrauer, J. Pernthaler, and C. von Mering. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Research*, 20(7):947–959, July 2010. doi: 10.1101/gr.104521.109. URL http://genome.cshlp.org/content/20/7/947.abstract.

J.-P. Changeux. Allostery and the Monod-Wyman-Changeux Model After 50 Years. *Annual review of biophysics*, 41:103–33, May 2011. doi: 10.1146/annurev-biophys-050511-102222.

J. M. Chase. Community assembly: when should history matter? *Oecologia*, 136(4):489–498, Aug. 2003. ISSN 1432-1939. doi: 10.1007/s00442-003-1311-7. URL https://doi.org/10.1007/s00442-003-1311-7.

J. M. Chase and J. A. Myers. Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1576):2351–2363, Aug. 2011. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2011.0063. URL https://royalsocietypublishing.org/doi/10.1098/rstb.2011.0063.

C. Chassagnole, N. Noisommit-Rizzi, J. W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of Escherichia coli. *Biotechnol Bioeng*, 79, 2002. doi: 10.1002/bit.10288. URL https://doi.org/10.1002/bit.10288.

H. C. Chiu, R. Levy, and E. Borenstein. Energent biosynthetic capacity in simple microbial communities. *PLOS computational biology*, 2014.

H.-H. Chou and C. Marx. Optimization of Gene Expression through Divergent Mutational Paths. *Cell Reports*, 1(2):133–140, Feb. 2012. ISSN 2211-1247. doi: 10.1016/j.celrep.2011.12.003. URL https://doi.org/10.1016/j.celrep.2011.12.003.

M. L. Coleman, M. B. Sullivan, A. C. Martiny, C. Steglich, K. Barry, E. F. DeLong, and S. W. Chisholm. Genomic Islands and the Ecology and Evolution of Prochlorococcus. *Science*, 311(5768):1768, Mar. 2006. doi: 10.1126/science.1122050. URL http://science.sciencemag.org/content/311/5768/1768.abstract.

F. L. Condamine, J. Rolland, and H. Morlon. Assessing the causes of diversification slowdowns: temperature-dependent and diversity-dependent models receive equivalent support. *Ecology Letters*, 22(11):1900–1912, Nov. 2019. ISSN 1461-023X. doi: 10.1111/ele.13382. URL https://doi.org/10.1111/ele.13382.

T. F. Cooper and R. E. Lenski. Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. *BMC evolutionary biology*, 10:11–11, Jan. 2010. ISSN 1471-2148. doi: 10.1186/1471-2148-10-11. URL https://pubmed.ncbi.nlm.nih.gov/20070898.

A. Cornish-Bowden. *Fundamentals of enzyme kinetics*. Butterworths, London ; Boston, 1979. ISBN 978-0-408-10617-7.

J. A. Coyne. Genetics and speciation. *Nature*, 355:511, Feb. 1992. URL http://dx.doi.org/10.1038/355511a0.

J. A. Coyne. *Why evolution is true*. Viking, New York (N.Y., 2010. ISBN 978-0-14-311664-6. OCLC: 1014527540.

J. A. Coyne and H. A. Orr. *Speciation*. Sinauer Associates, Sunderland, Mass, 2004. ISBN 978-0-87893-091-3 978-0-87893-089-0. OCLC: ocm55078441.

R. Daniel. The metagenomics of soil. *Nature Reviews Microbiology*, 3(6):470–478, June 2005. ISSN 1740-1526, 1740-1534. doi: 10.1038/nrmicro1160. URL http://www.nature.com/articles/nrmicro1160.

M. S. Datta, K. S. Korolev, I. Cvijovic, C. Dudley, and J. Gore. Range expansion promotes cooperation in an experimental microbial metapopulation. *Proceedings of the National Academy of Sciences*, 110(18):7354–7359, Apr. 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1217517110. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1217517110.

R. Dawkins. Extended Phenotype – But Not Too Extended. A Reply to Laland, Turner and Jablonka. *Biology and Philosophy*, 19(3):377–396, June 2004. ISSN 1572-8404. doi: 10.1023/B:BIPH.0000036180.14904.96. URL https://doi.org/10.1023/B:BIPH.0000036180.14904.96.

R. L. Day, K. N. Laland, and F. J. Odling-Smee. Rethinking Adaptation: The Niche-Construction Perspective. *Perspectives in Biology and Medicine*, 46(1):80–95, 2003.

K. de Queiroz. Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences*, 102(suppl 1):6600, May 2005. doi: 10.1073/pnas.0502030102. URL http://www.pnas.org/content/102/suppl_1/6600.abstract.

K. De Queiroz. Species Concepts and Species Delimitation. *Systematic Biology*, 56(6):879–886, Dec. 2007. ISSN 1063-5157. doi: 10.1080/10635150701701083. URL https://doi.org/10.1080/10635150701701083.

R. J. DeBerardinis and N. S. Chandel. Fundamentals of cancer metabolism. *Science Advances*, 2(5), May 2016. doi: 10.1126/sciadv.1600200. URL http://advances.sciencemag.org/content/2/5/e1600200.abstract.

M. DeJongh, K. Formsma, P. Boillot, J. Gould, M. Rycenga, and A. Best. Toward the automated generation of genome-scale metabolic networks in the SEED. *BMC bioinformatics*, 8:139–139, Apr. 2007. ISSN 1471-2105. doi: 10.1186/1471-2105-8-139. URL https://www.ncbi.nlm.nih.gov/pubmed/17462086.

O. Dias, M. Rocha, E. C. Ferreira, and I. Rocha. Reconstructing genome-scale metabolic models with merlin. *Nucleic Acids Research*, 43(8):3899–3910, Apr. 2015. ISSN 0305-1048. doi: 10.1093/nar/gkv294. URL https://doi.org/10.1093/nar/gkv294.

U. Dieckmann and M. Doebeli. On the origin of species by sympatric speciation. *Nature*, 400:354, July 1999a. URL http://dx.doi.org/10.1038/22521.

U. Dieckmann and M. Doebeli. On the origin of species by sympatric speciation. *Nature*, 400(6742):354–357, July 1999b. ISSN 1476-4687. doi: 10.1038/22521. URL https://doi.org/10.1038/22521.

F. Dini-Andreote, M. de Cássia Pereira e Silva, X. Triadó-Margarit, E. O. Casamayor, J. D. van Elsas, and J. F. Salles. Dynamics of bacterial community succession in a salt marsh chronosequence: evidences for temporal niche partitioning. *The ISME journal*, 8(10):1989–2001, Oct. 2014. ISSN 1751-7370. doi: 10.1038/ismej.2014.54. URL https://pubmed.ncbi.nlm.nih.gov/24739625.

M. Doebeli. A model for the evolutionary dynamics of cross-feeding polymorphisms in microorganisms. *Population Ecology*, 44(2):59–70, 2002. doi: 10.1007/s101440200008. URL http://dx.doi.org/10.1007/s101440200008.

G. D'Souza, S. Shitut, D. Preussger, G. Yousif, S. Waschina, and C. Kost. Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Natural Product Reports*, 35(5):455–488, 2018. ISSN 0265-0568. doi: 10.1039/C8NP00009C. URL http://dx.doi.org/10.1039/C8NP00009C.

A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7(1):74–74, Aug. 2013. doi: 10.1186/1752-0509-7-74. URL https://doi.org/10.1186/1752-0509-7-74.

J. S. Edwards and B. O. Palsson. Systems Properties of the *Haemophilus influenzae* Rd Metabolic Genotype. *Journal of Biological Chemistry*, 274(25):17410–17416, June 1999. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.274.25.17410. URL http://www.jbc.org/lookup/doi/10.1074/jbc.274.25.17410.

J. S. Edwards and B. O. Palsson. The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97 (10):5528, May 2000. doi: 10.1073/pnas.97.10.5528. URL http://www.pnas.org/content/97/10/5528.abstract.

J. S. Edwards, R. U. Ibarra, and B. O. Palsson. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat Biotechnol*, 19, 2001. doi: 10.1038/84379. URL https://doi.org/10.1038/84379.

B. C. Emerson and N. Kolm. Species diversity can drive speciation. *Nature*, 434(7036):1015–1017, Apr. 2005. ISSN 1476-4687. doi: 10.1038/nature03450. URL https://doi.org/10.1038/nature03450.

T. N. Enke, M. S. Datta, J. Schwartzman, N. Cermak, D. Schmitz, J. Barrere, A. Pascual-García, and O. X. Cordero. Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities. *Current Biology*, 29(9):1528–1535.e6, May 2019. ISSN 0960-9822. doi: 10.1016/j.cub.2019.03.047. URL http://www.sciencedirect.com/science/article/pii/S0960982219303458.

D. H. Erwin. Seeds of Diversity. *Science*, 308(5729):1752, June 2005. doi: 10.1126/science.1113416. URL http://science.sciencemag.org/content/308/5729/1752.abstract.

D. H. Erwin. Macroevolution of ecosystem engineering, niche construction and diversity. *Trends in Ecology & Evolution*, 23(6):304–310, June 2008. ISSN 0169-5347. doi: 10.1016/j.tree.2008.01.013. URL http://www.sciencedirect.com/science/article/pii/S0169534708001377.

S. Estrela and S. P. Brown. Metabolic and Demographic Feedbacks Shape the Emergent Spatial Structure and Function of Microbial Communities. *PLOS Computational Biology*, 9(12):e1003398, Dec. 2013. doi: 10.1371/journal.pcbi.1003398. URL https://doi.org/10.1371/journal.pcbi.1003398.

S. Estrela, C. H. Trisos, and S. P. Brown. From Metabolism to Ecology: Cross-Feeding Interactions Shape the Balance between Polymicrobial Conflict and Mutualism. *The American Naturalist*, 180 (5):566–576, Nov. 2012. ISSN 0003-0147. doi: 10.1086/667887. URL https://doi.org/10.1086/667887.

A. Eyre-Walker and P. D. Keightley. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8:610, Aug. 2007. URL https://doi.org/10.1038/nrg2146.

J. L. Feder, C. A. Chilcote, and G. L. Bush. Genetic differentiation between sympatric host races of the apple maggot fly Rhagoletis pomonella. *Nature*, 336:61, Nov. 1988. URL http://dx.doi.org/10.1038/336061a0.

A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. . Palsson. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3(1), 2007. doi: 10.1038/msb4100155. URL http://msb.embopress.org/content/3/1/121.

A. M. Feist, M. J. Herrgård, I. Thiele, J. L. Reed, and B. . Palsson. Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, 7(2):129–143, Feb. 2009. ISSN 1740-1526, 1740-1534. doi: 10.1038/nrmicro1949. URL http://www.nature.com/articles/nrmicro1949.

T. Fenchel and B. J. Finlay. The Ubiquity of Small Species: Patterns of Local and Global Diversity. *BioScience*, 54(8):777–784, Aug. 2004. ISSN 0006-3568. doi: 10.1641/0006-3568(2004)054[0777:TUOSSP]2.0.CO;2. URL https://doi.org/10.1641/0006-3568(2004)054[0777:TUOSSP]2.0.CO;2.

C. J. Fritzemeier, D. Hartleb, B. Szappanos, B. Papp, and M. J. Lercher. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLOS Computational Biology*, 13(4):e1005494, Apr. 2017. doi: 10.1371/journal.pcbi.1005494. URL https://doi.org/10.1371/journal.pcbi.1005494.

J. Förster, I. Famili, B. . Palsson, and J. Nielsen. Large-Scale Evaluation of *In Silico* Gene Deletions in *Saccharomyces cerevisiae*. *OMICS: A Journal of Integrative Biology*, 7(2):193–202, July 2003. ISSN 1536-2310, 1557-8100. doi: 10.1089/153623103322246584. URL http://www.liebertpub.com/doi/10.1089/153623103322246584.

T. Fukami. Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annual Review of Ecology, Evolution, and Systematics*, 46(1):1–23, Dec. 2015. ISSN 1543-592X. doi: 10.1146/annurev-ecolsys-110411-160340. URL https://doi.org/10.1146/annurev-ecolsys-110411-160340.

B. Futcher, G. I. Latter, P. Monardo, C. S. McLaughlin, and J. I. Garrels. A Sampling of the Yeast Proteome. *Molecular and Cellular Biology*, 19(11):7357, Nov. 1999. doi: 10.1128/MCB.19.11.7357. URL http://mcb.asm.org/content/19/11/7357.abstract.

G. F. Gause. *The struggle for existence*. Baltimore, MD: Williams & Wilkins., 1934.

A. H. Gentry. Tree species richness of upper Amazonian forests. *Proceedings of the National Academy of Sciences*, 85(1):156, Jan. 1988. doi: 10.1073/pnas.85.1.156. URL http://www.pnas.org/content/85/1/156.abstract.

P. Ghosh, J. Mondal, E. Ben-Jacob, and H. Levine. Mechanically-driven phase separation in a growing bacterial colony. *Proceedings of the National Academy of Sciences*, 112(17):E2166, Apr. 2015. doi: 10.1073/pnas.1504948112. URL http://www.pnas.org/content/112/17/E2166.abstract.

E. P. Gianchandani, A. K. Chavali, and J. A. Papin. The application of flux balance analysis in systems biology. *WIREs Systems Biology and Medicine*, 2(3):372–382, May 2010. ISSN 1939-5094. doi: 10.1002/wsbm.60. URL https://doi.org/10.1002/wsbm.60.

P. Gómez and A. Buckling. Real-time microbial adaptive diversification in soil. *Ecology Letters*, 16 (5):650–655, May 2013. ISSN 1461023X. doi: 10.1111/ele.12093. URL http://doi.wiley.com/10.1111/ele.12093.

A. Goldbeter and R. Lefever. Dissipative Structures for an Allosteric Model: Application to Glycolytic Oscillations. *Biophysical Journal*, 12(10):1302–1315, Oct. 1972. ISSN 0006-3495. doi: 10.1016/S0006-3495(72)86164-2. URL http://www.sciencedirect.com/science/article/pii/S0006349572861642.

J. E. Goldford, N. Lu, D. Bajić, S. Estrela, M. Tikhonov, A. Sanchez-Gorostiaga, D. Segrè, P. Mehta, and A. Sanchez. Emergent simplicity in microbial community assembly. *Science*, 361(6401):469, Aug. 2018. doi: 10.1126/science.aat1168. URL http://science.sciencemag.org/content/361/6401/469.abstract.

B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai. The dynamics of molecular evolution over 60,000 generations. *Nature*, 2017.

J. Gore, H. Youk, and A. van Oudenaarden. Snowdrift game dynamics and facultative cheating in yeast. *Nature*, 459(7244):253–256, May 2009. ISSN 1476-4687. doi: 10.1038/nature07921. URL https://www.ncbi.nlm.nih.gov/pubmed/19349960.

D. Greenbaum, C. Colangelo, K. Williams, and M. Gerstein. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*, 4(9):117, Aug. 2003a. ISSN 1474-760X. doi: 10.1186/gb-2003-4-9-117. URL https://doi.org/10.1186/gb-2003-4-9-117.

D. Greenbaum, C. Colangelo, K. Williams, and M. Gerstein. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome biology*, 4(9):117–117, 2003b. ISSN 1474-760X. doi: 10.1186/gb-2003-4-9-117. URL https://www.ncbi.nlm.nih.gov/pubmed/12952525.

T. Großkopf, J. Consuegra, J. Gaffé, J. C. Willison, R. E. Lenski, O. S. Soyer, and D. Schneider. Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. *BMC evolutionary biology*, 16(1):163, Aug. 2016. ISSN 1471-2148. doi: 10.1186/s12862-016-0733-x. URL http://europepmc.org/abstract/MED/27544664.

C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, and S. Y. Lee. Current status and applications of genome-scale metabolic models. *Genome Biology*, 20(1):121, June 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1730-3. URL https://doi.org/10.1186/s13059-019-1730-3.

I. Gudelj, R. E. Beardmore, S. S. Arkin, and R. C. MacLean. Constraints on microbial metabolism drive evolutionary diversification in homogeneous environments. *J Evol Biol*, 20, 2007. doi: 10.1111/j.1420-9101.2007.01376.x. URL https://doi.org/10.1111/j.1420-9101.2007.01376.x.

I. Gudelj, M. Kinnersley, P. Rashkov, K. Schmidt, and F. Rosenzweig. Stability of Cross-Feeding Polymorphisms in Microbial Communities. *PLOS Computational Biology*, 12(12):e1005269, Dec. 2016. doi: 10.1371/journal.pcbi.1005269. URL https://doi.org/10.1371/journal.pcbi.1005269.

S. Gudmundsson and I. Thiele. Computationally efficient flux variability analysis. *BMC Bioinformatics*, 11(1):489, Sept. 2010. ISSN 1471-2105. doi: 10.1186/1471-2105-11-489. URL https://doi.org/10.1186/1471-2105-11-489.

J. C. Guimaraes, M. Rocha, and A. P. Arkin. Transcript level and sequence determinants of protein abundance and noise in Escherichia coli. *Nucleic acids research*, 42(8):4791–4799, Apr. 2014. ISSN 1362-4962. doi: 10.1093/nar/gku126. URL https://www.ncbi.nlm.nih.gov/pubmed/24510099.

S. P. Gygi, Y. Rochon, B. R. Franza, and R. Aebersold. Correlation between Protein and mRNA Abundance in Yeast. *Molecular and Cellular Biology*, 19(3):1720, Mar. 1999. doi: 10.1128/MCB.19.3.1720. URL http://mcb.asm.org/content/19/3/1720.abstract.

C. A. Hanson, J. A. Fuhrman, M. C. Horner-Devine, and J. B. H. Martiny. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10(7):497–506, July 2012. ISSN 1740-1534. doi: 10.1038/nrmicro2795. URL https://doi.org/10.1038/nrmicro2795.

W. R. Harcombe, N. F. Delaney, N. Leiby, N. Klitgord, and C. J. Marx. The Ability of Flux Balance Analysis to Predict Evolution of Central Metabolism Scales with the Initial Distance to the Optimum. *PLoS Comput Biol*, 9(6):e1003091–e1003091, June 2013. doi: 10.1371/journal.pcbi.1003091. URL http://dx.doi.org/10.1371%2Fjournal.pcbi.1003091.

W. R. Harcombe, W. J. Riehl, I. Dukovski, B. R. Granger, A. Betts, A. H. Lang, G. Bonilla, A. Kar, N. Leiby, P. Mehta, C. J. Marx, and D. Segre. Metabolic resource allocation in individual microbes

determines ecosystem interactions and spatial dynamics. *Cell Reports*, 2014a. doi: 10.1016/j.celrep.2014.03.070.

W. R. Harcombe, W. J. Riehl, I. Dukovski, B. R. Granger, A. Betts, A. H. Lang, G. Bonilla, A. Kar, N. Leiby, P. Mehta, C. J. Marx, and D. Segre. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Reports*, 2014b. doi: 10.1016/j.celrep.2014.03.070.

G. Hardin. The Competitive Exclusion Principle. *Science*, 131(3409):1292, Apr. 1960. doi: 10.1126/science.131.3409.1292. URL http://science.sciencemag.org/content/131/3409/1292.abstract.

O. Hädicke and S. Klamt. EColiCore2: a reference network model of the central metabolism of Escherichia coli and relationships to its genome-scale parent model. *Scientific Reports*, 7(1):39647, Jan. 2017. ISSN 2045-2322. doi: 10.1038/srep39647. URL https://doi.org/10.1038/srep39647.

D. Healey, K. Axelrod, and J. Gore. Negative frequency-dependent interactions can underlie phenotypic heterogeneity in a clonal microbial population. *Molecular Systems Biology*, 12(8):877, Aug. 2016. ISSN 1744-4292. doi: 10.15252/msb.20167033. URL https://doi.org/10.15252/msb.20167033.

R. Heinrich and T. A. Rapoport. A Linear Steady-State Treatment of Enzymatic Chains. *European Journal of Biochemistry*, 42(1):89–95, Feb. 1974. ISSN 0014-2956. doi: 10.1111/j.1432-1033.1974.tb03318.x. URL https://doi.org/10.1111/j.1432-1033.1974.tb03318.x.

R. B. Helling, C. N. Vargas, and J. Adams. Evolution of Escherichia coli during growth in a constant environment. *Genetics*, 1987.

M. A. Hensen and T. J. Hanly. Dynamic flux balance analysis for synthetic microbial communities. *IET Syst. Biol.*, 2014.

B. Hess and A. Boiteux. Mechanism of Glycolytic Oscillation in Yeast, I. Aerobic and anaerobic growth conditions for obtaining glycolytic oscillation. *Hoppe-Seyler´s Zeitschrift für physiologische Chemie*, 349(2):1567–1574, Jan. 1968. ISSN 0018-4888. doi: 10.1515/bchm2.1968.349.2.1567. URL https://www.degruyter.com/view/j/bchm2.1968.349.issue-2/bchm2.1968.349.2.1567/bchm2.1968.349.2.1567.xml.

M. Higashi, G. Takimoto, and N. Yamamura. Sympatric speciation by sexual selection. *Nature*, 402:523, Dec. 1999. URL http://dx.doi.org/10.1038/990087.

H. Hillebrand. On the Generality of the Latitudinal Diversity Gradient. *The American Naturalist*, 163 (2):192–211, Feb. 2004. ISSN 0003-0147. doi: 10.1086/381004. URL https://doi.org/10.1086/381004.

M. J. A. v. Hoek and R. M. H. Merks. Emergence of microbial diversity due to cross-feeding interactions in a spatial model of gut microbial metabolism. *BMC Systems Biology*, 11(1):56, May 2017. ISSN 1752-0509. doi: 10.1186/s12918-017-0430-4. URL https://doi.org/10.1186/s12918-017-0430-4.

T. A. Hoek, K. Axelrod, T. Biancalani, E. A. Yurtsev, J. Liu, and J. Gore. Resource Availability Modulates the Cooperative and Competitive Nature of a Microbial Cross-Feeding Mutualism. *PLOS Biology*, 14(8):e1002540, Aug. 2016. doi: 10.1371/journal.pbio.1002540. URL https://doi.org/10.1371/journal.pbio.1002540.

Holmgren, M., Scheffer, M., and Huston, M. A. The interplay of facilitation and competition in plant communities. *Ecology*, 78:1966–1975, 1997. doi: 10.1890/0012-9658(1997)078[1966:TIOFAC]2.0.CO;2.

J. Hong and D. Gresham. Molecular Specificity, Convergence and Constraint Shape Adaptive Evolution in Nutrient-Poor Environments. *PLOS Genetics*, 10(1):e1004041, Jan. 2014. doi: 10.1371/journal.pgen.1004041. URL https://doi.org/10.1371/journal.pgen.1004041.

S.-R. Hosseini and A. Wagner. The potential for non-adaptive origins of evolutionary innovations in central carbon metabolism. *BMC Systems Biology*, 10(1):97, Oct. 2016. ISSN 1752-0509. doi: 10.1186/s12918-016-0343-7. URL https://doi.org/10.1186/s12918-016-0343-7.

S.-R. Hosseini, O. C. Martin, and A. Wagner. Phenotypic innovation through recombination in genome-scale metabolic networks. *Proceedings of the Royal Society B: Biological Sciences*, 283 (1839):20161536, Sept. 2016. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2016.1536. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2016.1536.

D. E. Hunt, L. A. David, D. Gevers, S. P. Preheim, E. J. Alm, and M. F. Polz. Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton. *Science*, 320(5879):1081, May 2008. doi: 10.1126/science.1157890. URL http://science.sciencemag.org/content/320/5879/1081.abstract.

G. E. Hutchinson. Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22(0):415–427, Jan. 1957. ISSN 0091-7451, 1943-4456. doi: 10.1101/SQB.1957.022.01.039. URL http://symposium.cshlp.org/cgi/doi/10.1101/SQB.1957.022.01.039.

G. E. Hutchinson. Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? *The American Naturalist*, 93(870):145–159, 1959. ISSN 00030147, 15375323. URL www.jstor.org/stable/2458768.

R. Ibarra, J. Edwards, and B. Palsson. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 2002.

T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood. Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. *Science*, 292(5518):929, May 2001. doi: 10.1126/science.292.5518.929. URL http://science.sciencemag.org/content/292/5518/929.abstract.

R. K. Jain, L. L. Munn, and D. Fukumura. Dissecting tumour pathophysiology using intravital microscopy. *Nature Reviews Cancer*, 2:266, Apr. 2002. URL http://dx.doi.org/10.1038/nrc778.

H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, Oct. 2000. ISSN 1476-4687. doi: 10.1038/35036627. URL https://doi.org/10.1038/35036627.

John C. Butcher. *Numerical Methods for Ordinary Differential Equations*. New York, second edition edition, 2003. ISBN 978-0-471-96758-3.

D. R. Johnson, F. Goldschmidt, E. E. Lilja, and M. Ackermann. Metabolic specialization and the assembly of microbial communities. *The Isme Journal*, 6:1985, May 2012. URL http://dx.doi.org/10.1038/ismej.2012.46.

K. A. Johnson and R. S. Goody. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry*, 50(39):8264–8269, Oct. 2011. ISSN 0006-2960. doi: 10.1021/bi201284u. URL https://doi.org/10.1021/bi201284u.

C. G. Jones, J. H. Lawton, and M. Shachak. Organisms as Ecosystem Engineers. In F. B. Samson and F. L. Knopf, editors, *Ecosystem Management: Selected Readings*, pages 130–147. Springer New York, New York, NY, 1996. ISBN 978-1-4612-4018-1. doi: 10.1007/978-1-4612-4018-1_14. URL https://doi.org/10.1007/978-1-4612-4018-1_14.

A. Jousset, N. Eisenhauer, M. Merker, N. Mouquet, and S. Scheu. High functional diversity stimulates diversification in experimental microbial communities. *Science Advances*, 2(6):e1600124, June 2016. doi: 10.1126/sciadv.1600124. URL http://advances.sciencemag.org/content/2/6/e1600124.abstract.

H. Kacser and J. A. Burns. The control of flux. *Symposia of the Society for Experimental Biology*, 27:65–104, 1973. ISSN 0081-1386. URL https://www.ncbi.nlm.nih.gov/pubmed/4148886.

M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1):D353–D361, Nov. 2016. ISSN 0305-1048. doi: 10.1093/nar/gkw1092. URL https://doi.org/10.1093/nar/gkw1092.

R. Kassen. The experimental evolution of specialists, generalists, and the maintenance of diversity. *Journal of Evolutionary Biology*, 15(2):173–190, Mar. 2002. ISSN 1420-9101. doi: 10.1046/j.1420-9101.2002.00377.x. URL http://dx.doi.org/10.1046/j.1420-9101.2002.00377.x.

R. Kassen and P. Rainey. The ecology and ggenetic of microbial diversity. *Annual Review of Microbiology*, 2004.

J. P. Keener and J. Sneyd. *Mathematical physiology*. Number 8 in Interdisciplinary applied mathematics. Springer, New York, NY, 2nd ed edition, 2009. ISBN 978-0-387-09419-9 978-0-387-75846-6 978-0-387-79387-0. OCLC: ocn298595247.

J. E. Keymer, M. A. Fuentes, and P. A. Marquet. Diversity emerging: from competitive exclusion to neutral coexistence in ecosystems. *Theoretical Ecology*, 5(3):457–463, Aug. 2012. ISSN 1874-1746. doi: 10.1007/s12080-011-0138-9. URL https://doi.org/10.1007/s12080-011-0138-9.

Z. A. King, J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. O. Palsson, and N. E. Lewis. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44(D1):D515–D522, 2016. doi: 10.1093/nar/gkv1049. URL +http://dx.doi.org/10.1093/nar/gkv1049.

M. A. Kinnersley, W. E. Holben, and F. Rosenzweig. E Unibus Plurum: Genomic Analysis of an Experimentally Evolved Polymorphism in Escherichia coli. *PLOS Genetics*, 5(11):e1000713, Nov. 2009. doi: 10.1371/journal.pgen.1000713. URL https://doi.org/10.1371/journal.pgen.1000713.

M. A. Kinnersley, J. Wenger, E. Kroll, J. Adams, G. Sherlock, and F. Rosenzweig. Ex Uno Plures: Clonal Reinforcement Drives Evolution of a Simple Microbial Community. *PLoS Genet*, 10(6):e1004430–e1004430, June 2014. doi: 10.1371/journal.pgen.1004430. URL http://dx.doi.org/10.1371%2Fjournal.pgen.1004430.

S. Klamt, U.-U. Haus, and F. Theis. Hypergraphs and Cellular Networks. *PLOS Computational Biology*, 5(5):e1000385, May 2009. doi: 10.1371/journal.pcbi.1000385. URL https://doi.org/10.1371/journal.pcbi.1000385.

N. Klitgord and D. Segrè. Ecosystems biology of microbial metabolism. *Curr Opin Biotechnol*, 22, 2011. doi: 10.1016/j.copbio.2011.04.018. URL https://doi.org/10.1016/j.copbio.2011.04.018.

A. Koeppel, E. B. Perry, J. Sikorski, D. Krizanc, A. Warner, D. M. Ward, A. P. Rooney, E. Brambilla, N. Connor, R. M. Ratcliff, E. Nevo, and F. M. Cohan. Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences*, 105(7):2504–2509, Feb. 2008. doi: 10.1073/pnas.0712205105. URL http://www.pnas.org/content/105/7/2504.abstract.

A. F. Koeppel, J. O. Wertheim, L. Barone, N. Gentile, D. Krizanc, and F. M. Cohan. Speedy speciation in a bacterial microcosm: new species can arise as frequently as adaptations within a species. *The ISME Journal*, 2013. doi: http://doi.org/10.1038/ismej.2013.3. URL http://doi.org/10.1038/ismej.2013.3.

A. S. Kondrashov and F. A. Kondrashov. Interactions among quantitative traits in the course of sympatric speciation. *Nature*, 400:351, July 1999. URL http://dx.doi.org/10.1038/22514.

D. E. Koshland, G. Némethy, and D. Filmer. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits*. *Biochemistry*, 5(1):365–385, Jan. 1966. ISSN 0006-2960. doi: 10.1021/bi00865a047. URL https://doi.org/10.1021/bi00865a047.

C. J. Krebs. *Ecology: The Experimental Analysis of Distribution and Abundance*. Pearson, Edinburgh Gate, 6. ed edition, 2014. ISBN 978-1-292-02627-5. OCLC: 862798675.

S. Kryazhimskiy, D. P. Rice, E. R. Jerison, and M. M. Desai. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, 344(6191):1519, June 2014. doi: 10.1126/science.1250939. URL http://science.sciencemag.org/content/344/6191/1519.abstract.

K. Laland, B. Matthews, and M. W. Feldman. An introduction to niche construction theory. *Evolutionary Ecology*, 30(2):191–202, Apr. 2016. ISSN 1573-8477. doi: 10.1007/s10682-016-9821-z. URL https://doi.org/10.1007/s10682-016-9821-z.

K. N. Laland, F. J. Odling-Smee, and M. W. Feldman. Evolutionary consequences of niche construction and their implications for ecology. *Proceedings of the National Academy of Sciences*, 96(18):10242–10247, Aug. 1999. doi: 10.1073/pnas.96.18.10242. URL http://www.pnas.org/content/96/18/10242.abstract.

A. Larhlimi, L. David, J. Selbig, and A. Bockmayr. F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC bioinformatics*, 13:57–57, Apr. 2012. ISSN 1471-2105. doi: 10.1186/1471-2105-13-57. URL https://www.ncbi.nlm.nih.gov/pubmed/22524245.

D. C. Laughlin, C. Joshi, P. M. van Bodegom, Z. A. Bastow, and P. Z. Fulé. A predictive model of community assembly that incorporates intraspecific trait variation. *Ecology Letters*, 15(11):1291–1299, Nov. 2012. ISSN 1461-023X. doi: 10.1111/j.1461-0248.2012.01852.x. URL https://doi.org/10.1111/j.1461-0248.2012.01852.x.

M. Le Gac, J. Plucain, T. Hindré, R. E. Lenski, and D. Schneider. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with Escherichia coli. *Proceedings of the National Academy of Sciences*, 109(24):9487, June 2012. doi: 10.1073/pnas.1207091109. URL http://www.pnas.org/content/109/24/9487.abstract.

A. L. Lehninger, D. L. Nelson, and M. M. Cox. *Lehninger principles of biochemistry*. W.H. Freeman, New York, 6th ed edition, 2013. ISBN 978-1-4292-3414-6. OCLC: ocn820352899.

R. E. Lenski. Dynamics of Interactions between Bacteria and Virulent Bacteriophage. In K. C. Marshall, editor, *Advances in Microbial Ecology*, volume 10, pages 1–44. Springer US, Boston, MA, 1988. ISBN 978-1-4684-5411-6 978-1-4684-5409-3. doi: 10.1007/978-1-4684-5409-3_1. URL http://link.springer.com/10.1007/978-1-4684-5409-3_1.

R. Levy and E. Borenstein. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences*, 110(31): 12804–12809, July 2013. doi: 10.1073/pnas.1300926110. URL http://www.pnas.org/content/110/31/12804.abstract.

N. E. Lewis, K. K. Hixson, T. M. Conrad, J. A. Lerman, P. Charusanti, and A. D. Polpitiya. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol*, 6, 2010. doi: 10.1038/msb.2010.47. URL https://doi.org/10.1038/msb.2010.47.

E. Litchman and C. A. Klausmeier. Trait-Based Community Ecology of Phytoplankton. *Annual Review of Ecology, Evolution, and Systematics*, 39(1):615–639, Oct. 2008. ISSN 1543-592X. doi: 10.1146/annurev.ecolsys.39.110707.173549. URL https://doi.org/10.1146/annurev.ecolsys.39.110707.173549.

K. Locey and J. Lennon. Scaling laws predict global microbial diversity. *PNAS*, 2016.

J. L. Lockwood, R. D. Powell, M. P. Nott, and S. L. Pimm. Assembling Ecological Communities in Time and Space. *Oikos*, 80(3):549–553, 1997. ISSN 00301299, 16000706. doi: 10.2307/3546628. URL www.jstor.org/stable/3546628.

A. J. Lotka. Contribution to the Theory of Periodic Reactions. *The Journal of Physical Chemistry*, 14(3):271–274, Mar. 1910. ISSN 0092-7325. doi: 10.1021/j150111a004. URL https://doi.org/10.1021/j150111a004.

S. Louca, P. M. Shih, M. W. Pennell, W. W. Fischer, L. W. Parfrey, and M. Doebeli. Bacterial diversification through geological time. *Nature Ecology & Evolution*, 2(9):1458–1467, Sept. 2018. ISSN 2397-334X. doi: 10.1038/s41559-018-0625-0. URL http://www.nature.com/articles/s41559-018-0625-0.

C. M. Loudon, B. Matthews, D. S. Sevilgen, and B. W. Ibelings. Experimental evidence that evolution by niche construction affects dissipative ecosystem dynamics. *Evolutionary Ecology*, 30(2):221–234, Apr. 2016. ISSN 1573-8477. doi: 10.1007/s10682-015-9802-7. URL https://doi.org/10.1007/s10682-015-9802-7.

C. A. Lozupone and R. Knight. Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 104(27):11436–11440, July 2007. ISSN 0027-8424. doi: 10.1073/pnas.0611525104. URL https://www.ncbi.nlm.nih.gov/pubmed/17592124.

T. R. Maarleveld, R. A. Khandelwal, B. G. Olivier, B. Teusink, and F. J. Bruggeman. Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnology journal*, 8(9):997–1008, Sept. 2013. ISSN 1860-7314. doi: 10.1002/biot.201200291. URL https://www.ncbi.nlm.nih.gov/pubmed/23893965.

R. C. MacLean, A. Dickson, and G. Bell. Resource competition and adaptive radiation in a microbial microcosm. *Ecology Letters*, 2005.

N. Madi, M. Vos, P. Legendre, and B. J. Shapiro. Diversity begets diversity in microbiomes. preprint, Evolutionary Biology, Apr. 2019. URL http://biorxiv.org/lookup/doi/10.1101/612739.

S. Magnúsdóttir, A. Heinken, L. Kutt, D. A. Ravcheev, E. Bauer, A. Noronha, K. Greenhalgh, C. Jäger, J. Baginska, P. Wilmes, R. M. T. Fleming, and I. Thiele. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, 35(1):81–89, Jan. 2017. ISSN 1546-1696. doi: 10.1038/nbt.3703. URL https://doi.org/10.1038/nbt.3703.

R. Mahadevan, J. S. Edwards, and F. J. Doyle. Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophys J*, 83, 2002. doi: 10.1016/S0006-3495(02)73903-9. URL https://doi.org/10.1016/S0006-3495(02)73903-9.

R. Maharjan, S. Seeto, L. Notley-McRobb, and T. Ferenci. Clonal Adaptive Radiation in a Constant Environment. *Science*, July 2006. doi: 10.1126/science.1129865. URL http://science.sciencemag.org/content/early/2006/07/06/science.1129865.abstract.

J. Martiny, B. Bohannan, J. Brown, R. Colwell, J. Fuhrman, J. Green, C. Horner-Devine, M. Kane, J. Krumins, C. Kuske, P. Morin, S. Naeem, L. Øvreås, a.-l. Reysenbach, V. Smith, and J. Staley. Microbial biogeography: putting microorganisms on the map. *Nature reviews. Microbiology*, 4: 102–12, Mar. 2006. doi: 10.1038/nrmicro1341.

B. Matthews, L. De Meester, C. G. Jones, B. W. Ibelings, T. J. Bouma, V. Nuutinen, J. van de Koppel, and J. Odling-Smee. Under niche construction: an operational bridge between ecology, evolution, and ecosystem science. *Ecological Monographs*, 84(2):245–263, May 2014. ISSN 1557-7015. doi: 10.1890/13-0953.1. URL http://dx.doi.org/10.1890/13-0953.1.

E. Mayr. *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press, Cambridge, Mass, 1st harvard university press pbk. ed edition, 1999. ISBN 978-0-674-86250-0.

B. Mcgill, B. Enquist, E. Weiher, and M. Westoby. Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, 21(4):178–185, Apr. 2006. ISSN 01695347. doi: 10.1016/j.tree.2006.02.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0169534706000334.

J. S. McKinnon and H. D. Rundle. Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution*, 17(10):480–488, Oct. 2002. ISSN 0169-5347. doi: 10.1016/S0169-5347(02)02579-X. URL http://dx.doi.org/10.1016/S0169-5347(02)02579-X.

C. P. McNally and E. Borenstein. Metabolic Model-Based Analysis of the Emergence of Bacterial Cross-Feeding through Extensive Gene Loss. *bioRxiv*, Jan. 2017. doi: 10.1101/180208. URL http://biorxiv.org/content/early/2017/08/23/180208.abstract.

L. McNally and S. P. Brown. Building the microbiome in health and disease: niche construction and social conflict in bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1675), Aug. 2015. doi: 10.1098/rstb.2014.0298. URL http://rstb.royalsocietypublishing.org/content/370/1675/20140298.abstract.

J. R. Meyer and R. Kassen. The effects of competition and predation on diversification in a model adaptive radiation. *Nature*, 446(7134):432–435, Mar. 2007. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature05599. URL http://www.nature.com/articles/nature05599.

J. R. Meyer, D. T. Dobias, S. J. Medina, L. Servilio, A. Gupta, and R. E. Lenski. Ecological speciation of bacteriophage lambda in allopatry and sympatry. *Science*, Nov. 2016. doi: 10.1126/science.aai8446. URL http://science.sciencemag.org/content/early/2016/11/21/science.aai8446.abstract.

R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38(suppl_1):D750–D753, Oct. 2009. ISSN 0305-1048. doi: 10.1093/nar/gkp889. URL https://doi.org/10.1093/nar/gkp889.

B. D. Mishler. The Morphological, Developmental, and Phylogenetic Basis of Species Concepts in Bryophytes. *The Bryologist*, 88(3):207, 1985. ISSN 00072745. doi: 10.2307/3243030. URL https://www.jstor.org/stable/3243030?origin=crossref.

S. Mitri, J. B. Xavier, and K. R. Foster. Social evolution in multispecies biofilms. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl 2(Suppl 2):10839–10846, June 2011. ISSN 1091-6490. doi: 10.1073/pnas.1100292108. URL https://www.ncbi.nlm.nih.gov/pubmed/21690380.

M. L. Mo, B. . Palsson, and M. J. Herrgård. Connecting extracellular metabolomic measurements to

intracellular flux states in yeast. *BMC Systems Biology*, 3(1):37, Mar. 2009. ISSN 1752-0509. doi: 10.1186/1752-0509-3-37. URL https://doi.org/10.1186/1752-0509-3-37.

C. Mora, D. P. Tittensor, S. Adl, A. G. B. Simpson, and B. Worm. How Many Species Are There on Earth and in the Ocean? *PLOS Biology*, 9(8):e1001127, Aug. 2011. doi: 10.1371/journal.pbio.1001127. URL https://doi.org/10.1371/journal.pbio.1001127.

R. Moreno-Sánchez, E. Saavedra, S. Rodríguez-Enríquez, and V. Olín-Sandoval. Metabolic control analysis: a tool for designing strategies to manipulate metabolic pathways. *Journal of biomedicine & biotechnology*, 2008:597913–597913, 2008. ISSN 1110-7251. doi: 10.1155/2008/597913. URL https://pubmed.ncbi.nlm.nih.gov/18629230.

M. Mori, M. Ponce-de León, J. Peretó, and F. Montero. Metabolic Complementation in Bacterial Communities: Necessary Conditions and Optimality. *Frontiers in Microbiology*, 7:1553, 2016. ISSN 1664-302X. doi: 10.3389/fmicb.2016.01553. URL https://www.frontiersin.org/article/10.3389/fmicb.2016.01553.

C. D. Nadell, K. R. Foster, and J. B. Xavier. Emergence of Spatial Structure in Cell Groups and the Evolution of Cooperation. *PLOS Computational Biology*, 6(3):e1000716, Mar. 2010. doi: 10.1371/journal.pcbi.1000716. URL https://doi.org/10.1371/journal.pcbi.1000716.

D. R. Nemergut, S. P. Anderson, C. C. Cleveland, A. P. Martin, A. E. Miller, A. Seimon, and S. K. Schmidt. Microbial Community Succession in an Unvegetated, Recently Deglaciated Soil. *Microbial Ecology*, 53(1):110–122, Jan. 2007. ISSN 1432-184X. doi: 10.1007/s00248-006-9144-7. URL https://doi.org/10.1007/s00248-006-9144-7.

D. R. Nemergut, S. K. Schmidt, T. Fukami, S. P. O'Neill, T. M. Bilinski, L. F. Stanish, J. E. Knelman, J. L. Darcy, R. C. Lynch, P. Wickey, and S. Ferrenberg. Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews*, 77(3):342–356, Sept. 2013. ISSN 1092-2172. doi: 10.1128/MMBR.00051-12. URL http://mmbr.asm.org/cgi/doi/10.1128/MMBR.00051-12.

R. A. Notebaart, B. Teusink, R. J. Siezen, and B. Papp. Co-Regulation of Metabolic Genes Is Better Explained by Flux Coupling Than by Network Distance. *PLOS Computational Biology*, 4(1):e26, Jan. 2008. doi: 10.1371/journal.pcbi.0040026. URL https://doi.org/10.1371/journal.pcbi.0040026.

F. J. Odling-Smee, K. N. Laland, and M. W. Feldman. *Niche construction: the neglected process in evolution*. Number no. 37 in Monographs in population biology. Princeton University Press, Princeton, 2003. ISBN 978-0-691-04438-5 978-0-691-04437-8.

J. D. Orth, I. Thiele, and B. O. Palsson. What is flux balance analysis? *Nature biotechnology*, 2010. doi: 10.1038/nbt.1614.

J. D. Orth, T. M. Conrad, J. Na, J. A. Lerman, H. Nam, A. M. Feist, and B. O. Palsson. A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. *Mol Syst Biol.*, 2011.

N. Paczia, A. Nilgen, T. Lehmann, J. Gätgens, W. Wiechert, and S. Noack. Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microbial Cell Factories*, 11(1):122, Sept. 2012. ISSN 1475-2859. doi: 10.1186/1475-2859-11-122. URL https://doi.org/10.1186/1475-2859-11-122.

M. W. Palmer and T. A. Maurer. Does Diversity Beget Diversity? A Case Study of Crops and Weeds. *Journal of Vegetation Science*, 8(2):235–240, 1997. ISSN 11009233, 16541103. doi: 10.2307/3237352. URL www.jstor.org/stable/3237352.

S. Pande and C. Kost. Bacterial Unculturability and the Formation of Intercellular Metabolic Networks. *Trends in Microbiology*, 25(5):349–361, May 2017. ISSN 0966-842X. doi: 10.1016/j.tim.2017.02.015. URL http://www.sciencedirect.com/science/article/pii/S0966842X17300525.

J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson. Comparison of network-based pathway analysis methods. *Trends in Biotechnology*, 22(8):400–405, Aug. 2004. ISSN 0167-7799. doi: 10.1016/j.tibtech.2004.06.010. URL http://www.sciencedirect.com/science/article/pii/S0167779904001751.

G. A. Pavlopoulos, M. Secrier, C. N. Moschopoulos, T. G. Soldatos, S. Kossida, J. Aerts, R. Schneider, and P. G. Bagos. Using graph theory to analyze biological networks. *BioData Mining*, 4(1):10, Apr. 2011. ISSN 1756-0381. doi: 10.1186/1756-0381-4-10. URL https://doi.org/10.1186/1756-0381-4-10.

E. Pennisi. The Man Who Bottled Evolution. *Science*, 342(6160):790–793, Nov. 2013. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.342.6160.790. URL https://www.sciencemag.org/lookup/doi/10.1126/science.342.6160.790.

T. Pfeiffer and S. Bonhoeffer. Evolution of cross-feeding in microbial populations. *Am Nat*, 2004.

T. Pfeiffer, S. Schuster, and S. Bonhoeffer. Cooperation and competition in the evolution of ATP-producing pathways. *Science*, 292, 2001. doi: 10.1126/science.1058079. URL https://doi.org/10.1126/science.1058079.

P. Pharkya, E. V. Nikolaev, and C. D. Maranas. Review of the BRENDA Database. *Metabolic Engineering*, 5(2):71–73, Apr. 2003. ISSN 1096-7176. doi: 10.1016/S1096-7176(03)00008-9. URL http://www.sciencedirect.com/science/article/pii/S1096717603000089.

E. R. Pianka. Latitudinal Gradients in Species Diversity: A Review of Concepts. *The American Naturalist*, 100(910):33–46, Jan. 1966. ISSN 0003-0147. doi: 10.1086/282398. URL https://doi.org/10.1086/282398.

J. Plucain, T. Hindré, M. Le Gac, O. Tenaillon, S. Cruveiller, C. Médigue, N. Leiby, W. R. Harcombe, C. J. Marx, R. E. Lenski, and D. Schneider. Epistasis and Allele Specificity in the Emergence of a Stable Polymorphism in Escherichia coli. *Science*, 343(6177):1366–1369, 2014. doi: 10.1126/science.1248688. URL http://science.sciencemag.org/content/343/6177/1366.

E. Porcher, O. Tenaillon, and B. Godelle. From metabolism to polymorphism in bacterial populations: A theoretical study. *Evolution*, 55(11):2181–2193, Nov. 2001. ISSN 1558-5646. doi: 10.1111/j.0014-3820.2001.tb00734.x. URL http://dx.doi.org/10.1111/j.0014-3820.2001.tb00734.x.

N. D. Price, I. Famili, D. A. Beard, and B. . Palsson. Extreme Pathways and Kirchhoff's Second Law. *Biophysical Journal*, 83(5):2879–2882, Nov. 2002a. ISSN 0006-3495. doi: 10.1016/S0006-3495(02)75297-1. URL http://www.sciencedirect.com/science/article/pii/S0006349502752971.

N. D. Price, J. A. Papin, and B. O. Palsson. Determination of Redundancy and Systems Properties of the Metabolic Network of Helicobacter pylori Using Genome-Scale Extreme Pathway Analysis. *Genome Research*, 12(5):760–769, Apr. 2002b. ISSN 1088-9051. doi: 10.1101/gr.218002. URL http://www.genome.org/cgi/doi/10.1101/gr.218002.

N. D. Price, J. L. Reed, and B. . Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*, 2(11):886–897, Nov. 2004. ISSN 1740-1526, 1740-1534. doi: 10.1038/nrmicro1023. URL http://www.nature.com/articles/nrmicro1023.

T. D. Price, D. M. Hooper, C. D. Buchanan, U. S. Johansson, D. T. Tietze, P. Alström, U. Olsson, M. Ghosh-Harihar, F. Ishtiaq, S. K. Gupta, J. Martens, B. Harr, P. Singh, and D. Mohan. Niche filling slows the diversification of Himalayan songbirds. *Nature*, 509(7499):222–225, May 2014. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature13272. URL http://www.nature.com/articles/nature13272.

I. Rabbers, J. H. van Heerden, N. Nordholt, H. Bachmann, B. Teusink, and F. J. Bruggeman. Metabolism at evolutionary optimal states. *Metabolites*, 2015.

D. L. Rabosky and A. H. Hurlbert. Species Richness at Continental Scales Is Dominated by Ecological

Limits. *The American Naturalist*, 185(5):572–583, May 2015. ISSN 0003-0147. doi: 10.1086/680850. URL https://doi.org/10.1086/680850.

P. B. Rainey and M. Travisano. Adaptive radiation in a heterogeneous environment. *Nature*, 1998.

K. Raman and N. Chandra. Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics*, 10(4):435–449, July 2009. ISSN 1467-5463. doi: 10.1093/bib/bbp011. URL http://dx.doi.org/10.1093/bib/bbp011.

D. Raubenheimer and S. Simpson. Integrating nutrition: a geometrical approach. *Entomologia Experimentalis et Applicata*, 91(1):67–82, Apr. 1999. ISSN 1570-7458. doi: 10.1046/j.1570-7458.1999.00467.x. URL http://dx.doi.org/10.1046/j.1570-7458.1999.00467.x.

E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási. Hierarchical Organization of Modularity in Metabolic Networks. *Science*, 297(5586):1551, Aug. 2002. doi: 10.1126/science.1073374. URL http://science.sciencemag.org/content/297/5586/1551.abstract.

K. D. Rawls, B. V. Dougherty, E. M. Blais, E. Stancliffe, G. L. Kolling, K. Vinnakota, V. R. Pannala, A. Wallqvist, and J. A. Papin. A simplified metabolic network reconstruction to promote understanding and development of flux balance analysis tools. *Computers in Biology and Medicine*, 105:64–71, Feb. 2019. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2018.12.010. URL http://www.sciencedirect.com/science/article/pii/S0010482518304086.

Q. Ren, B. Henes, M. Fairhead, and L. Thöny-Meyer. High level production of tyrosinase in recombinant Escherichia coli. *BMC Biotechnology*, 13(1):18, Dec. 2013. ISSN 1472-6750. doi: 10.1186/1472-6750-13-18. URL https://bmcbiotechnol.biomedcentral.com/articles/10.1186/1472-6750-13-18.

C. Rocabert, C. Knibbe, J. Consuegra, D. Schneider, and G. Beslon. Beware batch culture: Seasonality and niche construction predicted to favor bacterial adaptive diversification. *PLOS Computational Biology*, 13(3):e1005459, Mar. 2017. doi: 10.1371/journal.pcbi.1005459. URL https://doi.org/10.1371/journal.pcbi.1005459.

J. F. M. Rodrigues and A. Wagner. Evolutionary Plasticity and Innovations in Complex Metabolic Reaction Networks. *Plos Computational Biology*, 2009.

R. F. Rosenzweig, R. R. Sharp, D. S. Treves, and J. Adams. Microbial evolution in a simple unstructured environment: genetic differentiation in Escherichia coli. *Genetics*, 1994.

R. Rosselló-Mora and R. Amann. The species concept for prokaryotes. *FEMS Microbiology Reviews*, 25(1):39–67, Jan. 2001. ISSN 0168-6445. doi: 10.1111/j.1574-6976.2001.tb00571.x. URL https://doi.org/10.1111/j.1574-6976.2001.tb00571.x.

D. E. Rozen and R. E. Lenski. Long-term experimental evolution in Escherichia coli. VII. Dynamics of a balanced polymorphism. *Am Nat*, 2000a.

D. E. Rozen and R. E. Lenski. Long-term experimental evolution in Escherichia coli. VIII. Dynamics of a balanced polymorphism. *Am Nat*, 155, 2000b. doi: 10.1086/303299. URL https://doi.org/10.1086/303299.

D. E. Rozen, D. Schneider, and R. E. Lenski. Long-Term Experimental Evolution in Escherichia coli. XIII. Phylogenetic History of a Balanced Polymorphism. *Journal of Molecular Evolution*, 61(2):171–180, Aug. 2005. ISSN 0022-2844, 1432-1432. doi: 10.1007/s00239-004-0322-2. URL http://link.springer.com/10.1007/s00239-004-0322-2.

A. Samal and O. C. Martin. Randomizing Genome-Scale Metabolic Networks. *PLOS ONE*, 6(7): e22295, July 2011. doi: 10.1371/journal.pone.0022295. URL https://doi.org/10.1371/journal.pone.0022295.

A. Samal, J. F. M. Rodrigues, J. Jost, O. C. Martin, and A. Wagner. Genotype networks in metabolic reaction spaces. *BMC Systems Biology*, 2010.

M. San Roman and A. Wagner. An enormous potential for niche construction through bacterial cross-feeding in a homogeneous environment. *PLOS Computational Biology*, 14(7):e1006340, July 2018. doi: 10.1371/journal.pcbi.1006340. URL https://doi.org/10.1371/journal.pcbi.1006340.

M. San Román, H. Cancela, and L. Acerenza. Source and regulation of flux variability in Escherichia coli. *BMC Systems Biology*, 8(1):67, June 2014. ISSN 1752-0509. doi: 10.1186/1752-0509-8-67. URL https://doi.org/10.1186/1752-0509-8-67.

T. E. Sandberg, C. J. Lloyd, B. O. Palsson, and A. M. Feist. Laboratory Evolution to Alternating Substrate Environments Yields Distinct Phenotypic and Genetic Adaptive Strategies. *Applied and Environmental Microbiology*, 83(13):e00410–17, July 2017. doi: 10.1128/AEM.00410-17. URL http://aem.asm.org/content/83/13/e00410-17.abstract.

H. L. Sanders. Marine Benthic Diversity: A Comparative Study. *The American Naturalist*, 102(925): 243–282, 1968. ISSN 00030147, 15375323. URL www.jstor.org/stable/2459027.

M. Scheer, A. Grote, A. Chang, I. Schomburg, C. Munaretto, M. Rother, C. Söhngen, M. Stelzer, J. Thiele, and D. Schomburg. BRENDA, the enzyme information system in 2011. *Nucleic Acids Res*, 2011.

J. Schellenberger, J. O. Park, T. M. Conrad, and B. . Palsson. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11(1):213,

Apr. 2010. ISSN 1471-2105. doi: 10.1186/1471-2105-11-213. URL https://doi.org/10.1186/1471-2105-11-213.

C. H. Schilling, S. Schuster, B. O. Palsson, and R. Heinrich. Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-genomic Era. *Biotechnology Progress*, 15(3):296–303, Jan. 1999. ISSN 8756-7938. doi: 10.1021/bp990048k. URL https://doi.org/10.1021/bp990048k.

D. Schluter and M. W. Pennell. Speciation gradients and the distribution of biodiversity. *Nature*, 546: 48, May 2017. URL https://doi.org/10.1038/nature22897.

R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology*, 3(1):119, Jan. 2007. ISSN 1744-4292. doi: 10.1038/msb4100162. URL https://doi.org/10.1038/msb4100162.

R. Schuster and H.-G. Holzhütter. Use of Mathematical Models for Predicting the Metabolic Effect of Large-Scale Enzyme Activity Alterations. *European Journal of Biochemistry*, 229(2):403–418, Apr. 1995. ISSN 0014-2956. doi: 10.1111/j.1432-1033.1995.0403k.x. URL https://doi.org/10.1111/j.1432-1033.1995.0403k.x.

D. Segrè, D. Vitkup, and G. M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112, Nov. 2002. doi: 10.1073/pnas.232349399. URL http://www.pnas.org/content/99/23/15112.abstract.

M. Serres and M. Riley. MultiFun, a Multifunctional Classification Scheme for Escherichia coli K-12 Gene Products. *Microbial & Comparative Genomics*, 5(4):205–222, Jan. 2000. ISSN 1090-6592. doi: 10.1089/omi.1.2000.5.205. URL https://doi.org/10.1089/omi.1.2000.5.205.

G. Sezonov, D. Joseleau-Petit, and R. D'Ari. Escherichia coli Physiology in Luria-Bertani Broth. *J. Bacteriol*, 2007. doi: 10.1128/JB.01368-07.

T. Shlomi, O. Berkman, and E. Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7695, May 2005. doi: 10.1073/pnas.0406346102. URL http://www.pnas.org/content/102/21/7695.abstract.

J. Sikorski and E. Nevo. Adaptation and incipient sympatric speciation of Bacillus simplex under microclimatic contrast at "Evolution Canyons" I and II, Israel. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44):15924–15929, Nov. 2005. doi: 10.1073/pnas.0507944102. URL http://www.pnas.org/content/102/44/15924.abstract.

J. M. Smith. Sympatric Speciation. *The American Naturalist*, 100(916):637–650, 1966. ISSN 00030147, 15375323. URL http://www.jstor.org/stable/2459301.

R. R. Stein, V. Bucci, N. C. Toussaint, C. G. Buffie, G. Rätsch, E. G. Pamer, C. Sander, and J. B. Xavier. Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. *PLOS Computational Biology*, 9(12):e1003388, Dec. 2013. doi: 10.1371/journal.pcbi.1003388. URL https://doi.org/10.1371/journal.pcbi.1003388.

J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193, Nov. 2002. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature01166. URL http://www.nature.com/articles/nature01166.

S. Stolyar, S. V. Dien, K. L. Hillesland, N. Pinel, T. J. Lie, J. A. Leigh, and D. A. Stahl. Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology*, 2007.

S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. d'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, Tara Oceans coordinators, C. Bowler, C. de Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, P. Bork, E. Boss, C. Bowler, M. Follows, L. Karp-Boss, U. Krzic, E. G. Reynaud, C. Sardet, M. Sieracki, and D. Velayoudon. Structure and function of the global ocean microbiome. *Science*, 348(6237):1261359–1261359, May 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1261359. URL https://www.sciencemag.org/lookup/doi/10.1126/science.1261359.

K. Takemoto. Current understanding of the formation and adaptation of metabolic systems based on network theory. *Metabolites*, 2(3):429–457, July 2012. ISSN 2218-1989. doi: 10.3390/metabo2030429. URL https://www.ncbi.nlm.nih.gov/pubmed/24957641.

O. Tenaillon, A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long, and B. S. Gaut. The Molecular Diversity of Adaptive Convergence. *Science*, 335(6067):457, Jan. 2012. doi: 10.1126/science.1212986. URL http://science.sciencemag.org/content/335/6067/457.abstract.

M. P. Thakur and A. J. Wright. Environmental Filtering, Niche Construction, and Trait Variability: The Missing Discussion. *Trends in Ecology & Evolution*, 32(12):884–886, Dec. 2017. ISSN 0169-

5347. doi: 10.1016/j.tree.2017.09.014. URL http://www.sciencedirect.com/science/article/pii/S0169534717302574.

The Earth Microbiome Project Consortium, L. R. Thompson, J. G. Sanders, D. McDonald, A. Amir, J. Ladau, K. J. Locey, R. J. Prill, A. Tripathi, S. M. Gibbons, G. Ackermann, J. A. Navas-Molina, S. Janssen, E. Kopylova, Y. Vázquez-Baeza, A. González, J. T. Morton, S. Mirarab, Z. Zech Xu, L. Jiang, M. F. Haroon, J. Kanbar, Q. Zhu, S. Jin Song, T. Kosciolek, N. A. Bokulich, J. Lefler, C. J. Brislawn, G. Humphrey, S. M. Owens, J. Hampton-Marcell, D. Berg-Lyons, V. McKenzie, N. Fierer, J. A. Fuhrman, A. Clauset, R. L. Stevens, A. Shade, K. S. Pollard, K. D. Goodwin, J. K. Jansson, J. A. Gilbert, and R. Knight. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551(7681):457–463, Nov. 2017. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature24621. URL http://www.nature.com/articles/nature24621.

The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature*, 486(7402):215–221, June 2012. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature11209. URL http://www.nature.com/articles/nature11209.

D. Tilman. *Resource competition and community structure*. Number 17 in Monographs in population biology. Princeton University Press, Princeton, N.J, 1982. ISBN 978-0-691-08301-8 978-0-691-08302-5.

E. Toprak, A. Veres, J.-B. Michel, R. Chait, D. L. Hartl, and R. Kishony. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature Genetics*, 44:101, Dec. 2011. URL https://doi.org/10.1038/ng.1034.

M. Travisano and P. B. Rainey. Studies of adaptive radiation using model microbial systems. *American Naturalist*, 2000.

D. S. Treves, S. Manning, and J. Adams. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of Escherichia coli. *Molecular Biology and Evolution*, 15(7):789–797, 1998. URL http://mbe.oxfordjournals.org/content/15/7/789.abstract.

S.-D. Tsen, S.-C. Lai, C.-P. Pang, J.-I. Lee, and T. Wilson. Chemostat Selection of anEscherichia coliMutant Containing Permease with Enhanced Lactose Affinity. *Biochemical and Biophysical Research Communications*, 224(2):351–357, July 1996. ISSN 0006-291X. doi: 10.1006/bbrc.1996.1032. URL http://www.sciencedirect.com/science/article/pii/S0006291X96910323.

J. Turner. *The Extended Organism: The Physiology of Animal-Built Structures*. Harvard University Press, 2009. ISBN 978-0-674-04449-4. URL https://books.google.ch/books?id=wapnxKbb23MC.

E. Tzamali, P. Poirazi, I. G. Tollis, and M. Reczko. A computational exploration of bacterial metabolic diversity identifying metabolic interactions and growth-efficient strain communities. *BMC Systems Biology*, 2011.

M. Vallejo-Marín, R. J. A. Buggs, A. M. Cooley, and J. R. Puzey. Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species Mimulus peregrinus. *Evolution*, 69(6):1487–1500, June 2015. ISSN 0014-3820. doi: 10.1111/evo.12678. URL https://doi.org/10.1111/evo.12678.

B. Van den Bergh, T. Swings, M. Fauvart, and J. Michiels. Experimental Design, Population Dynamics, and Diversity in Microbial Experimental Evolution. *Microbiology and Molecular Biology Reviews*, 82 (3):e00008–18, Sept. 2018. doi: 10.1128/MMBR.00008-18. URL http://mmbr.asm.org/content/82/3/e00008-18.abstract.

A. Varma and B. O. Palsson. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology*, 12(10):994–998, Oct. 1994a. ISSN 1546-1696. doi: 10.1038/nbt1094-994. URL https://doi.org/10.1038/nbt1094-994.

A. Varma and B. O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl Environ Microbiol.*, 1994b.

M. Vellend. *The theory of ecological communities*. Monographs in population biology. Princeton University Press, Princeton, 2016. ISBN 978-0-691-16484-7.

M. Vellend and A. Agrawal. Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology*, 85(2):183–206, 2010. ISSN 00335770, 15397718. doi: 10.1086/652373. URL www.jstor.org/stable/10.1086/652373.

D. Vitkup, P. Kharchenko, and A. Wagner. Influence of metabolic network structure and function on enzyme evolution. *Genome Biology*, 7(5):R39, May 2006. ISSN 1474-760X. doi: 10.1186/gb-2006-7-5-r39. URL https://doi.org/10.1186/gb-2006-7-5-r39.

C. Vogel and E. M. Marcotte. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13:227, Mar. 2012. URL https://doi.org/10.1038/nrg3185.

A. Wagner and D. A. Fell. The small world inside large metabolic networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1478):1803–1810, Sept. 2001. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2001.1711. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2001.1711.

H. Wang, S. Marcišauskas, B. J. Sánchez, I. Domenzain, D. Hermansson, R. Agren, J. Nielsen, and E. J. Kerkhoven. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. *PLOS Computational Biology*, 14(10):e1006541, Oct. 2018. doi: 10.1371/journal.pcbi.1006541. URL https://doi.org/10.1371/journal.pcbi.1006541.

O. Warburg. On respiratory impairment in cancer cells. *Science (New York, N.Y.)*, 124(3215):269–270, Aug. 1956. ISSN 0036-8075. URL http://europepmc.org/abstract/MED/13351639.

M. P. Washburn, A. Koller, G. Oshiro, R. R. Ulaszek, D. Plouffe, C. Deciu, E. Winzeler, and J. R. Yates, 3rd. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, 100(6):3107–3112, Mar. 2003. ISSN 0027-8424. doi: 10.1073/pnas.0634629100. URL https://www.ncbi.nlm.nih.gov/pubmed/12626741.

D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684): 440–442, June 1998. ISSN 1476-4687. doi: 10.1038/30918. URL https://doi.org/10.1038/30918.

J. Whitfield. Is Everything Everywhere? *Science*, 310(5750):960, Nov. 2005. doi: 10.1126/science.310.5750.960. URL http://science.sciencemag.org/content/310/5750/960.abstract.

R. H. Whittaker. Evolution and Measurement of Species Diversity. *Taxon*, 21(2/3):213–251, 1972. ISSN 00400262. doi: 10.2307/1218190. URL http://www.jstor.org/stable/1218190.

S. Widder, R. J. Allen, T. Pfeiffer, T. P. Curtis, C. Wiuf, W. T. Sloan, O. X. Cordero, S. P. Brown, B. Momeni, W. Shou, H. Kettle, H. J. Flint, A. F. Haas, B. Laroche, J.-U. Kreft, P. B. Rainey, S. Freilich, S. Schuster, K. Milferstedt, J. R. van der Meer, T. Grosskopf, J. Huisman, A. Free, C. Picioreanu, C. Quince, I. Klapper, S. Labarthe, B. F. Smets, H. Wang, O. S. Soyer, and Isaac Newton Institute Fellows. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME Journal*, 10(11):2557–2568, Nov. 2016. ISSN 1751-7370. doi: 10.1038/ismej.2016.45. URL https://doi.org/10.1038/ismej.2016.45.

T. E. Wood, N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences*, 106(33):13875–13879, 2009. ISSN 0027-8424. doi: 10.1073/pnas.0811575106. URL https://www.pnas.org/content/106/33/13875.

T. K. Wood, S. J. Knabel, and B. W. Kwan. Bacterial persister cell formation and dormancy. *Applied and environmental microbiology*, 79(23):7116–7121, Dec. 2013. ISSN 1098-5336. doi: 10.1128/AEM.02636-13. URL https://pubmed.ncbi.nlm.nih.gov/24038684.

L. Wu, M. R. Mashego, A. M. Proell, J. L. Vinke, C. Ras, J. van Dam, W. A. van Winden, W. M. van Gulik, and J. J. Heijnen. In vivo kinetics of primary metabolism in Saccharomyces cerevisiae studied through prolonged chemostat cultivation. *Metabolic Engineering*, 8(2):160–171, Mar. 2006. ISSN 1096-7176. doi: 10.1016/j.ymben.2005.09.005. URL http://www.sciencedirect.com/science/article/pii/S1096717605000777.

B. Xu, M. Jahic, G. Blomsten, and S.-O. Enfors. Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with Escherichia coli. *Applied Microbiology and Biotechnology*, 51(5):564–571, May 1999. ISSN 1432-0614. doi: 10.1007/s002530051433. URL https://doi.org/10.1007/s002530051433.

X. Yi and A. M. Dean. Bounded population sizes, fluctuating selection and the tempo and mode of coexistence. *Proceedings of the National Academy of Sciences*, 110(42):16945, Oct. 2013. doi: 10.1073/pnas.1309830110. URL http://www.pnas.org/content/110/42/16945.abstract.

You-Kwan Oh, Bernhard O. Palsson, Sung M. Park, Christophe H. Schilling, and Radhakrishnan Mahadevan. Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data. *Journal of Biological Chemistry*, Sept. 2007. doi: 10.1074/jbc.M703759200.

A. Zelezniak, S. Andrejev, O. Ponomarova, D. R. Mende, P. Bork, and K. R. Patil. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences*, 112(20):6449–6454, May 2015. doi: 10.1073/pnas.1421834112. URL http://www.pnas.org/content/112/20/6449.abstract.

K. Zhuang, M. Izallalen, P. Mouser, H. Richter, C. Risso, R. Mahadevan, and D. R. Lovley. Genome-scale dynamic modeling of the competition between Rhodoferax and Geobacter in anoxic subsurface environments. *The ISME Journal*, 5(2):305–316, Feb. 2011. ISSN 1751-7362, 1751-7370. doi: 10.1038/ismej.2010.117. URL http://www.nature.com/articles/ismej2010117.

A. R. Zomorrodi and C. D. Maranas. OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities. *PLoS Comput Biol*, 8(2):e1002363–e1002363, Feb. 2012. doi: 10.1371/journal.pcbi.1002363. URL http://dx.doi.org/10.1371%2Fjournal.pcbi.1002363.

A. R. Zomorrodi and D. Segrè. Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities. *Nature communications*, 8(1):1563, Nov. 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-01407-5. URL http://europepmc.org/abstract/MED/29146901.

Curriculum Vitae

Personal details

Name	Magdalena San Román Rincón
Nationality	Uruguayan
Date of birth	6 March 1988
Contact	magdalena.sanroman@ieu.uzh.ch

Education

03/2015-today	**PhD student in Evolutionary Biology** *University of Zurich*, Switzerland Supervisor: Prof. Andreas Wagner
09/2010-12/2013	**Master of Science in Biology, Biophysics** *University of the Republic*, Uruguay Supervisor: Prof. Luis Acerenza, Prof. Héctor Cancela Graduated with distinction
03/2005-07/2010	**Bachelor in Biochemistry** *University of the Republic*, Uruguay

Research and Work experience

03/2014-02/2015	**Research fellow at Prof. Etchebehere lab** *Institute Clemente Estable*, Uruguay
03/2014-02/2015	**Teacher Grade 2, Biophysics Section** University of the Republic, Uruguay
August 2012	**Short-term visit at Cedersund Lab** *Linköping University*, Sweden
03/2010-03/2013	**Research fellow at Prof. Acerenza lab** *University of the Republic*, Uruguay
03/2009-02/2014	**Teacher Grade 1, Biophysics Section** University of the Republic, Uruguay

Scientific publications

M. San Román, A. Wagner. Acetate and glycerol are not uniquely suited for the evolution of cross-feeding in *E. coli*. *Submitted* (2020)

J. Wenzel, A. Ferraz Júnior, **M. San Román**, A. Cabezas, C. Etchebehere. Unraveling the causes of low H2 yield in dark fermentation of cheese whey using Clostridium and Enterobacter pure cultures. *Submitted* (2019)

M. San Román, A. Wagner. An enormous potential for niche construction through bacterial cross-feeding in a homogeneous environment. *PLOS Computational Biology* (2018)

M. San Román, H. Cancela, L. Acerenza. Source and regulation of flux variability in *Escherichia coli*. *BMC Systems Biology* (2014)

Teaching, Supervising and Organizational Experience

2019	**Teaching Assistant** in 'BIO352 Principles of Evolution'. University of Zürich (UZH)
2018	**Teaching Assistant** in 'BIO 249 Research Internship in Quantitative and System Biology'. UZH
2017	**Lab rotation supervision** of student Cheng Li. UZH
2016-today	**Teaching Assistant** in 'BIO 351 Principles of Evolution'. UZH
2016	**Teaching Assistant** in 'BIO 134 Programming in Biology'. UZH
2015-today	**Teaching Assistant** in 'BIO 334 Practical Bioinformatics'. UZH
2013	**Co-organizer** of course 'Solving exercises using computational tool'. University of the Republic (UdelaR)
2013	**Co-organizer** of '2nd Conference +Biophysics' Institute Pasteur (IP) and UdelaR
2012	**Teaching Assistant** in 'A Systems Biology Approach to Understanding Mechanisms of Organismal Evolution'. IP and UdelaR
2011	**Teaching Assistant** in 'Yeast Systems Biology'. IP
2010-2015	**Teaching Assistant** in 'Biophysics'. UdelaR
2010 and 2012	**Teaching Assistant** in 'Systems Biology'. UdelaR

Dissemination to general public

Interview in Uruguayan radio show 'Sobre Ciencia' (2018)

M. San Román. A Change, An Opportunity: How Bacteria Can Affect A Homogeneous Environment. *Science Trends* (2018)

M. San Román. En el afán de descubrirlo todo: sobre Carlos Alberto Cal. *Uruguay Ciencia* (2012)

Languages and Skills

Spanish (native), English (advanced), German (intermediate), Portuguese (intermediate)

Python, Matlab, R, Mathematica, LaTeX